

Colloque

MESURES DE L'INTERNET

CMI'03

12, 13 et 14 mai 2003

Hôtel Westminster, Nice (France)

Président :

Eric GUICHARD, INRIA Sophia-Antipolis

Introduction

Le but de ce colloque est d'approfondir le dialogue entre l'ensemble des experts pour qui l'internet est un objet d'étude, indépendamment de leur approche (sciences exactes ou humaines). Des spécialistes des pratiques intellectuelles et des technologues seront aussi présents, afin de proposer une réflexion sur la notion de mesure et sur l'objet internet.

Il s'agit de favoriser une intercompréhension des problématiques, des méthodes et des analyses, à notre avis essentielle à la recherche, mais aussi fort utile aux entreprises, institutions politiques et médiatiques, et au grand public : l'internet est à la fois un réseau technique et une production humaine, et aborder l'un sans l'autre, ou aborder les deux sans rigueur génère plus de bruit et de confusion que de résultats scientifiques utiles à l'humanité.

Une première expérience de ce type, très fructueuse, avait déjà été réalisée en décembre 1999 à l'École normale supérieure, sur le thème des usages de l'internet

Ce colloque constitue donc la seconde du genre, et, au vu de la très grande qualité des intervenants, répond sans nul doute à un fort besoin des chercheurs de partager leurs travaux avec des confrères de disciplines différentes, et en retour, de s'imprégner des résultats de ces collègues.

Cette manifestation témoigne de la puissance d'une interdisciplinarité motivée par une demande précise des acteurs de la recherche, bien au fait de leurs terrains et de leurs problématiques : elle s'avère alors concrète, fédératrice et prometteuse.

Éric Guichard

Comité Scientifique

Président : Éric Guichard (INRIA)
Daniel Andler (ENS, études cognitives)
Richard Arena (UNSA, économie)
François Baccelli (INRIA ENS, usages et statistiques)
Éric Brousseau (Paris-X, économie)
Fulvio Caccia (Canadiens en Europe, littérature)
Claire Charbit (ENST, économie)
Henri Desbois (Paris-X, cybergéographie)
Émilie Devriendt (ENS et Paris-IV, linguistique)
Jack Goody (Cambridge, anthropologie)
Jacques Lajoie (UQÀM, psychologie)
Yves Lechevallier (INRIA, statistiques et fouille de données)
Christian Licoppe (FT, usages et économie)
Paul Mathias (Henri IV, philosophie des techniques)
Frédéric Moatty (sociologie, CNRS CEE)
Jean-Pierre Nadal (CNRS ENS, systèmes complexes)
Serge Proulx (UQÀM, usages et sociologie)
Philippe Rygiel (Paris-I, histoire et pratiques intellectuelles)
Jacques-Philippe Saint-Gérard (Clermont-Ferrand et CNRS, linguistique)
Brigitte Trousse (INRIA, usages et parcours)
Bernard Victorri (CNRS ENS, linguistique)

Comité d'organisation

Il est composé de membres de l'équipe AXIS et du service des colloques de l'INRIA Sophia-Antipolis : Eric Guichard (responsable), Sophie Honnorat, Dany Sergeant, Brigitte Trousse, Marie-Hélène Zeitoun ainsi que des membres du service SEMIR.

Se sont associés de façon ponctuelle des membres de l'association Les Canadiens en Europe et de l'équipe Réseaux, Savoirs & Territoires (ENS).

Remerciements

Ceux-ci s'adressent tout d'abord à Suzanne Beauchamp (association Les Canadiens en Europe) et à Michel Cosnard (Directeur de l'unité de recherche Sophia Antipolis de l'INRIA), qui ont très vite et très fermement manifesté leur soutien total à cette manifestation.

Et aussi à l'ensemble des institutions et structures qui ont soutenu financièrement le colloque :

L'équipe AXIS, l'action concertée incitative GRID, l'association Les Canadiens en Europe, le Conseil général des Alpes Maritimes, le Conseil régional Provence Alpes Côte d'Azur, l'équipe Réseaux, Savoirs & Territoires de l'ENS, le groupement de recherche TICS du CNRS, le Laboratoire des Usages de Sophia Antipolis, le ministère de la Recherche et des Nouvelles Technologies.

Aux membres du comité scientifique, dont les contacts et les conseils ont été fort précieux, et qui ont eu la lourde tâche de sélectionner les propositions de communication.

Enfin, un grand merci à Sophie Honnorat, à Dany Sergeant et à Catherine Lefebvre, qui n'ont jamais compté leur temps ni leur énergie, et sans lesquelles ce colloque n'aurait pu se tenir.

L'appel à contribution

1. Écriture, informatique et internet

L'internet est depuis plusieurs années un objet d'étude à part entière. Il témoigne d'un processus d'informatisation engagé depuis une cinquantaine d'années, qui correspond à une transformation de la technique d'écriture : les ordinateurs travaillent essentiellement sur les systèmes graphiques (listes, tableaux de nombres, etc.) avec ces mêmes systèmes graphiques (code, programme, etc.). Il s'ensuit un accroissement de la variété et du nombre des objets écrits, du fait de la numérisation généralisée du texte, du son, de l'image, qui, tous, peuvent être manipulés. L'internet parachève ce phénomène en facilitant le repérage et l'appropriation de ces textes, en même temps que ses protocoles en produisent d'autres : les traces des pratiques individuelles ou agrégées.

Ces dernières années, le succès de la combinaison entre les réseaux physiques et les activités humaines a beaucoup surpris, comme en témoigne la récente bulle spéculative autour de l'internet. Mais il n'a pas étonné les scientifiques (toutes spécialités confondues), qui cherchent à optimiser leurs pratiques intellectuelles : l'écriture est le pivot de leurs activités, que ce soit dans le cadre de la recherche ou de la transmission des connaissances, et ils savent qu'une amélioration de leur instrument de travail génère souvent de puissants enjeux.

2. Disciplines impliquées

On conçoit donc que l'internet génère des recherches au sein de disciplines variées, et appelle à un débat entre ces diverses disciplines. Mais l'explicitation de la relation entre l'écriture, la pensée et la raison mérite d'être approfondie par les anthropologues et les philosophes. Ils contribueront notamment à l'acquisition d'une posture technologique, au sens fort du terme, ce qui aidera à la définition rigoureuse de concepts, préalables à des mesures originales et efficaces.

Pour analyser des usages ou imaginer des régulations, les sociologues et les économistes sont de plus en plus tentés par l'analyse des faits et conduites électroniques des personnes, en utilisant par exemple les archives fournies par des moteurs de recherche, des caches, des routeurs. Ces analyses quantitatives et textuelles apportent des résultats exceptionnels (tant sur le plan méthodologique que sur celui de la mise en évidence de pratiques non imaginées), à condition de travailler sur de gros volumes, et de ne pas négliger les phénomènes rares. Bien entendu, ce type de recherche n'exclut pas, bien au contraire, une confrontation avec des mesures ethnographiques.

Le cas est aussi flagrant avec la linguistique, qui retrouve indirectement une nouvelle vigueur avec le développement des moteurs de recherche : la langue écrite, qui se prête déjà naturellement à la formalisation probabiliste, invite, sous sa forme numérisée, à des formes de comptage qui vont jusqu'à la représentation géométrique d'espaces sémantiques. Les « lois intellectuelles du langage » se donnent à voir, tant grâce à la numérisation des dictionnaires anciens que grâce aux mises en correspondance de millions de sites web. Les outils linguistiques accompagnent de plus en plus fréquemment le texte, et transforment la démarche scientifique de l'ensemble des chercheurs en sciences humaines, quand les moteurs de recherche finissent par infléchir la notion de qualité scientifique avec leurs algorithmes de repérage de la renommée des chercheurs.

De nombreux ingénieurs, informaticiens, physiciens et mathématiciens travaillent aussi sur les archives électroniques de l'internet : flux IP de routeurs, échanges au sein d'un réseau national, etc. Leurs finalités sont variées : modélisation, contrôle, mise en évidence des lois physiques et de la topologie de l'internet..., le plus souvent à des fins d'optimisation de ce réseau « technique » en perpétuelle expansion. Mais là encore, la dynamique humaine est telle que ces formes de perfectionnement passent aussi par des mesures d'usages et des dispositifs d'aide, à la navigation et à la recherche, par exemple. Outre de nouveaux travaux dans les champs de la combinatoire, de l'algorithmique, de la théorie des graphes..., on remarque un affinement des modèles statistiques (rendu possible par les possibilités des machines à traiter l'écrit que sont les ordinateurs et la présence de données massives et variées). Ce qui conduit à deux types de constats particulièrement féconds : a) les questions de méthode (classification, illusion d'exhaustivité) et d'échantillonnage sont réactualisées ; b) les lois de puissance s'imposent, et il n'est pas étonnant que ces distributions soient aussi celles qu'ont découvertes les spécialistes des usages et les linguistes.

Cette convergence des mesures appliquées à l'homme, à ses pratiques d'échanges, et aux instruments qui font les réseaux incite à comprendre l'activité des machines comme une somme d'actions produites par les hommes, et à prévoir l'évolution des usages, comme cela a été réalisé dans le cas de l'essor du peer-to-peer. Réciproquement, les découvertes statistiques invitent à proposer des modèles de l'interaction sociale plus sophistiqués que ceux communément déduits de la loi de Gauss.

Enfin, les représentations graphiques issues des mesures de l'internet apparaissent comme une forme ultime de production écrite destinées à la synthèse ou à la conviction, tout en utilisant au mieux les potentialités de l'écriture contemporaine (cartes animées, graphiques réactifs, etc.). Cette manifestation de la réflexivité de l'écriture a évidemment des incidences sur d'autres représentations, plus sociales, comme celles qui se construisent autour des notions de territoire et d'identités : les cartes de l'internet (locales, globales, géographiques, mentales), parfois réalisées sans intention géographique, produisent du territoire, et donc infléchissent les représentations collectives, non seulement de l'internet, mais plus généralement de la planète, de soi et des « autres ». On retrouve une étroite association entre le cartographe et l'ingénieur, souvent euphémisée, néanmoins attestée depuis le XVIIe siècle.

3. Retombées

La demande est de plus en plus forte (chez les universitaires, les industriels et les responsables politiques) pour donner un cadre conceptuel à l'internet. Une telle réflexion « technologique », associée à des mesures physiques interprétées en termes de pratiques sociales, aide à optimiser les prévisions en matière d'usages collectifs (pour imaginer des produits, dimensionner des ressources, comprendre des engouements ou des résistances). Par ailleurs, les gestionnaires de réseaux ont un grand besoin de systèmes de visualisation de leurs trafics (sécurité, optimisation, prévision, etc.), lesquels s'intègrent parfois dans des productions culturelles.

Enfin, le dialogue interdisciplinaire profitera beaucoup de cet échange approfondi entre les spécialistes des mesures physiques et des mesures sociales.

RÉSUMÉS
DES
COMMUNICATIONS

Métrologie des réseaux de l'internet : principales actions et impact sur les évolutions technologiques

Philippe Owezarski

Chargé de recherches au CNRS
LAAS-CNRS
7, Avenue du Colonel Roche
31077 Toulouse cedex 4

owe@laas.fr

Cette conférence propose une introduction à la métrologie, d'un point de vue technique, et compte montrer comment les observations faites sur le trafic internet permettent d'orienter les recherches autour de la conception et de l'ingénierie des réseaux du futur (les NGN : Next Generation Networks).

Nous présenterons ici un tour d'horizon du domaine de la métrologie, en commençant par introduire sa problématique, notamment dans l'internet, et en dressant l'état de l'art des projets passés et actuels les plus significatifs, et leurs principales contributions. Nous décrirons successivement les techniques de mesures *actives* qui consistent à émettre des paquets sur le réseau et à effectuer des mesures sur la transmission de ces paquets sondes, et la métrologie *passive* qui consiste à observer le trafic qui transite en un ou plusieurs points du réseau au cours du temps, ou à conserver une trace plus ou moins fine de ce trafic. Dans les deux cas, les données obtenues doivent ensuite être analysées pour en tirer des informations sur les performances du réseau, sur la QoS obtenue par les flux, sur les matrices de trafic, sur la topologie du réseau, etc. Nous insistons sur le fait que tous les premiers résultats viennent des États-Unis, et qu'à quelques rares exceptions, l'Europe, la France et le reste du monde sont très peu impliqués dans ce type de mesures. Toutefois, depuis un peu plus d'un an, un projet national de recherche en métrologie a été lancé sous l'égide du RNRT.

Ce projet, intitulé METROPOLIS, a déjà débouché sur plusieurs résultats, qui contredisent les croyances collectives en matière de métrologie. Ainsi l'évolution du trafic de l'internet est-elle illustrée par des comparaisons faites entre ce qui a été observé de 2000 à 2003. Ces comparaisons portent sur la répartition du trafic par protocoles ou par applications ; notamment l'impact des nouvelles applications sur les caractéristiques du trafic est largement abordé. En fait, à partir des nouvelles caractéristiques de ce trafic, on découvre que les architectures protocolaires de l'internet ne sont plus adaptées aux nouveaux besoins ; par ailleurs ces caractéristiques du trafic observé vont orienter les recherches futures en métrologie. A terme, c'est la façon de concevoir le réseau et ses mécanismes qui est remise en cause.

D'autre part, même si on se cantonne à des considérations techniques, on est conduit à engager des recherches pluridisciplinaires dès que l'on aborde la métrologie des réseaux et du trafic internet. Afin de mener des études prospectives sur toutes les thématiques concernées par la métrologie et lancer des actions fédératives autour de ces domaines, le département *Société et Techniques de l'Information et de la Communication* (STIC) du CNRS, par l'intermédiaire de l'action du Réseau Thématique Pluridisciplinaire (RTP) «réseau de Communication», a démarré une *Action Spécifique* (AS) chargée de «débroussailler» le

domaine et de lui faire des recommandations quant aux actions à conduire. Les premières conclusions de cette AS seront décrites à la fin de notre communication.

Les dispositifs de construction de l'internaute par la mesure d'audience : une approche sociologique

Josiane Jouët

Professeur en sociologie de la communication

Institut Français de Presse
Université Panthéon-Assas, Paris II

jjouet@club-internet.fr

Notre communication se propose de questionner la notion d'audience appliquée à internet d'une part, et de mettre en perspective les apports et les limites de cette mesure d'audience d'autre part. Elle vise aussi à distinguer la représentation de l'internaute par la mesure d'audience de l'approche de l'internaute construit par la sociologie des usages.

La notion d'audience d'internet paraît de fait bien éloignée de celle des médias de masse traditionnels et cela à plusieurs titres. D'abord l'architecture technique d'internet repose sur une autre logique car on sort du modèle de la diffusion au profit d'un modèle de connexion. Nous ne sommes donc plus en présence d'une offre construite pour rejoindre simultanément de vastes auditoires car les internautes sont investis dans une activité de communication qui bouleverse l'espace-temps des audiences médiatiques. Par ailleurs, internet est un pluri-média qui donne l'accès à une multitude de services. C'est aussi un hyper-média qui combine les ressources de l'écrit, de l'image et du son et qui assigne à l'utilisateur une posture de communication interactive fort distincte de la posture de réception des médias de masse. Surtout, le média internet, parce qu'il offre une multitude d'applications et de services, rend impossible une mesure d'audience exhaustive. La différence avec les médias de masse est donc ici majeure et une vaste partie d'internet échappe de fait à toute mesure.

Néanmoins, la démocratisation d'internet et son glissement vers le modèle éditorial repérable dans la profusion de sites s'adressant au «grand public» justifie, du moins en partie, l'emprunt de la notion d'audience. Pourtant la mesure d'audience d'internet n'est pas un simple calque des mesures d'audience des médias car elle se greffe aussi d'emblée sur deux autres logiques : la logique de mesure du trafic des télécommunications et la logique de numérisation de l'informatique. Ainsi la mesure d'audience d'internet est d'abord et de façon massive une *mesure d'audience sur internet*.

Toutefois si internet est le média qui *a priori* peut le mieux connaître son audience, il est aussi le média le plus complexe à mesurer. Les outils de mesure d'audience et d'analyse deviennent de plus en plus pointus et ciblés autour de la représentation de l'internaute comme consommateur. La puissance des logiciels de capture, le raffinement des outils de traitement de données, la finesse des modèles d'analyse statistique produisent une évaluation en continu des activités sur internet. Néanmoins, ces dispositifs comportent des limites. Le marketing devient étroitement imbriqué aux mesures d'audience qui ne sont pas tant conçues pour la connaissance des publics que comme outils stratégiques. L'internaute mesuré devient un

produit vendu aux annonceurs ; il ne s'agit plus seulement ici de publics cibles comme pour les médias de masse, mais aussi de consommateurs individuels. En outre, la quantification excessive risque de ne plus produire que des modèles formels, des segmentations de publics, des stéréotypes de comportements, des typologies hâtives et des procédures de catégorisation. Par ailleurs, le caractère performatif des mesures agit non seulement sur l'offre éditoriale qui se moule sur un modèle de l'internaute calibré par les dispositifs, mais il imprègne aussi les discours médiatiques véhiculés autour des résultats d'audience.

Aussi les enjeux des dispositifs de mesure ne se bornent-ils pas aux questions soulevées par les modes de représentation de l'internaute, mais soulèvent-ils le débat sur la façon dont ils contribuent aussi peu à peu et en profondeur à la construction sociale du média.

Application des techniques de fouille de données aux logs :

état de l'art sur le Web Usage Mining

Doru Tanasa, doctorant
Brigitte Trousse, responsable scientifique de l'équipe Axis
Florent Masseglia, chercheur

INRIA Sophia Antipolis, Equipe AxIS
2004 route des lucioles BP 93, 06902 Sophia Antipolis

(Doru.Tanasa, Brigitte.Trousse, Florent.Masseglia)@sophia.inria.fr

La fouille de données (ou *data mining*) est un domaine de recherche encore assez jeune. Les techniques qui en sont issues sont motivées par le besoin des propriétaires de grandes masses de données «de faire parler» ces données. La question étant généralement «est-ce que mes données cachent quelque chose que je ne peux pas découvrir en raison de leur grande taille?». Motivée par des besoins de type marketing, la fouille de données cherche aujourd'hui à savoir si l'on peut identifier des schémas nouveaux valides, potentiellement utiles et surtout compréhensibles et utilisables. Aujourd'hui, nos capacités à stocker, exploiter, véhiculer et échanger des informations massives se sont largement accrues. Du groupe financier au centre de recherches médicales, en passant par la compagnie de télécommunications, l'information circule et s'empile à un rythme qui ne faiblit pas. Les moyens créant les besoins, cette nouvelle caractéristique des bases de données offre aujourd'hui des possibilités que les outils traditionnels, par la nature de leurs fonctionnalités, ne permettent pas d'exploiter.

Le cas est flagrant avec l'internet. Compte tenu de la rapidité à laquelle les fichiers *logs* (enregistrant les actions sur un site web) accumulent les données, les questions de la fouille de données sur de tels fichiers se sont rapidement posées. De cette application des techniques de fouille sur les données du web est apparu un domaine de recherche à part entière : le Web Mining.

Le Web Mining recouvre trois aspects : le *Web Content Mining*, le *Web Structure Mining* et le *Web Usage Mining*. La première approche, comme son nom l'indique, s'intéresse au contenu « propre » des pages web, la seconde à la structure et à l'organisation de ces pages. Le *Web Usage Mining* vise en revanche à repérer des motifs comportementaux chez les utilisateurs, à partir d'un ou plusieurs serveurs web.

Dans cette communication, nous proposons de faire le point sur les techniques de *web usage Mining* actuelles, expliquant à la fois les motivations des techniques engagées mais aussi les fondements de ces techniques. Nous expliquerons quels sont les paramètres pris en compte par le domaine du *Web Usage Mining* : les *logs*, mais aussi la structure du site ou encore les informations disponibles sur le profil utilisateur. Nous veillerons donc à clarifier les techniques, en nous appuyant sur des exemples (liés au web ou à d'autres domaines) et à présenter les mesures à prendre en compte, pour pouvoir prétendre à l'extraction de connaissances dans le contexte particulier du web.

Pour une bibliographie sur le Web Usage Mining voir : <http://www-sop.inria.fr/axis/wum/biblio.html>

Adaptive On-Line Page Importance Computation

Serge Abiteboul
INRIA et Xyleme

Mihai Preda
Xyleme

Grégory Cobena
INRIA
Gregory.Cobena@inria.fr

All the pages on the web do not have the same importance. For example, *Le Louvre* homepage is more important than an unknown person's homepage. Page importance information is very valuable. It is used by search engines to display results in the order of page importance (see Google). It is also useful for guiding the refreshing and discovery of pages: important pages should be refreshed more often and when crawling for new pages, important pages have to be fetched first.

Following some ideas of Kleinberg (1999), Page and Brin proposed a notion of page importance based on the link structure of the web. This was then used by Google with a remarkable success. Intuitively, a page is important if there are many important pages pointing to it. This corresponds, for instance, to the intuition of importance for research articles: a paper is important if it is referenced by many other important papers.

Formally, page importance, or PageRank is defined as the fix-point of a matrix equation.

We present a formal model of the web as a graph and we discuss the existence of a unique value of page importance, the algorithms to compute this value, and the modifications that may be applied to the simpler graph model of the web.

The computation of page importance in a huge dynamic graph has recently attracted a lot of attention because of the web. Previous algorithms computed it off-line and required the use of a lot of extra CPU as well as disk resources (e.g. to store, maintain and read the link matrix). We introduce a new algorithm OPIC that works on-line, and uses much less resources. In particular, it does not require storing the link matrix. It is on-line in that it continuously refines its estimate of page importance while the web/graph is visited. Thus it can be used to focus crawling to the most interesting pages. We prove the correctness of OPIC.

We present Adaptive OPIC that also works on-line but adapts dynamically to changes of the web. A variant of this algorithm is now used by Xyleme. We report on experiments with synthetic data. In particular, we study the convergence and adaptiveness of the algorithms for various scheduling strategies for the pages to visit. We also report on experiments based on crawls of significant portions of the web.

References:

- 1) “**Computing web page importance without storing the graph of the web**”(extended abstract), *IEEE-CS Data Engineering Bulletin*, Volume 25, 2002.
- 2) "**Adaptive On-Line Page Importance Computation**", *The Twelfth International World Wide Web Conference*, 2003, Budapest (to appear)

Quelques mesures du web comme corpus d'usages linguistiques

Russon Wooldridge

Professeur

University of Toronto
wulfric@chass.utoronto.ca

Nous présentons ici des **mesures quantitatives et qualitatives du web considéré comme un corpus d'usages linguistiques**. Si en anglais le mot *corpus* a le sens très répandu de «corpus linguistique» à cause de la place importante occupée par la *corpus linguistics* dans le monde anglophone et anglicisant de pays comme le Royaume-Uni, les pays scandinaves et les Pays-Bas, il en va autrement dans les pays francophones où la linguistique de corpus existe relativement peu (on note surtout quelques travaux sur la langue orale). L'étude du web comme corpus d'usages linguistiques, peu répandue encore dans le domaine de l'anglais, serait quasiment inexistante en français. Pourtant le web représente un corpus d'usages contemporains, dynamique donc essentiellement synchronique, un corpus tout fait, qui offre un terrain d'étude riche et mérite un examen sérieux.

Cette communication présente quelques modèles d'investigation et d'analyse, ainsi que quelques mesures des phénomènes étudiés.

Voici les phénomènes linguistiques observés dans ce bref tour d'horizon : les variantes et la productivité lexicale. Variantes d'expression, variantes géo-linguistiques, variantes morphologiques, confusions réelles ou feintes ; productivité formelle et sémantique d'un lexème, productivité paradigmatique de matrices syntagmatiques. Suffisance, dans la plupart des cas, du moteur de recherche Google ; suppléance, dans certains cas, par AltaVista. Quelques complexités de recherche : besoin de l'opérateur ET NON dans quelques cas ambigus ; nécessité de tenir compte de la variation *oe* / *ø*, dont la correspondance est transparente pour l'œil humain, opaque pour le moteur de recherche. Les mesures faites ici varient entre l'exact et le flou, mais elles ont toujours une valeur relative certaine.

Certains des exemples choisis illustrent l'intérêt socio-linguistique de l'exploration du web comme corpus d'usages linguistiques : comparaisons géo-linguistiques (cf. *courriel* vs. *email*, etc.) ; *esprit de corps* vs. *esprit d'équipe* ; distortions quantitatives provoquées par les sites pornographiques.

Dans plusieurs cas, nous confrontons l'usage observable dans le web avec les déclarations normatives du dictionnaire. Dans certains cas, les occurrences en ligne viennent enrichir et compléter ce qu'offre le dictionnaire (ex. *enfirouaper*) ; dans d'autres cas, l'usage réel entre en contradiction avec le dictionnaire, qui n'enregistre que ce qu'il juge sanctionné par la norme. C'est le cas, par exemple, de l'adjectif *chou* que le dictionnaire (*Petit Robert* et *TLFI*) déclare être invariable, mais que les occurrences du web montrent clairement variable, donnant au pluriel *chou*, *chous* ou *choux*.

Pour une zone qui se situerait entre le très fréquent (plusieurs centaines de milliers de résultats) et le très particulier (idiolectes) et dont les unités sont observables du point de vue de leur forme (syntagmatique vs. polysémie), le web peut offrir des instantanés mesurables de plusieurs types de phénomènes linguistiques, quitte à raffiner les résultats par des analyses de détail. C'est le seul corpus étendu qui ne demande pas une construction coûteuse en termes de temps et d'argent. Corpus peu théorisé donc et non la chasse gardée de spécialistes ou d'équipes de chercheurs en quête de subventions importantes. Corpus virtuel qui est transformé en corpus concret par l'utilisateur se servant intelligemment de moteurs de recherche. Pour la première fois, quiconque s'intéresse à l'usage réel de la langue dispose d'un champ d'observation de valeur scientifique autrement plus satisfaisant et complet que les analyses ponctuelles et les recueils de perles des auteurs de livres ou de chroniques.

L'avenir du Web-corpus est prometteur. En perpétuelle expansion (par ex. *navigateur(s)* = 9 793 occurrences en décembre 1997, 804 000 en mars 2003), il offre un terrain d'exploration toujours plus riche à l'observateur de la langue, que celui-ci se contente d'instantanés approximatifs ou s'attarde à développer des agrandissements à définition variable.

Analyse sémantique par ordinateur des courriels

de la ressource de cybermentorat Academos

Jacques Lajoie, professeur

Catherine Légaré, doctorante, Département de psychologie

Pierre Plante, Lucie Dumas, André Plante, Analystes en informatique, Centre ATO, faculté
des sciences humaines

Université du Québec à Montréal

lajoie.jacques@uqam.ca, clégare@academos.qc.ca, plante.pierre@uqam.ca,
dumas.lucie@uqam.ca, plante.andre@uqam.ca

L'évaluation scientifique des communautés virtuelles bénéficie du fait que toutes les communications textuelles laissent une trace. Cependant, la quantité gigantesque d'information peut devenir un obstacle aux méthodes traditionnelles d'analyse qualitative des textes. Celles-ci exigent habituellement un codage manuel des données et la catégorisation des thèmes de discussion se fait *a priori* par l'expérimentateur, ce qui crée une source importante de subjectivité. Le caractère scientifique de ces analyses provient alors d'une validation indépendante des analyses par des juges. Cette approche devient très onéreuse en temps lorsque plusieurs milliers de messages doivent être analysés. La présente recherche explore l'usage d'un nouvel outil d'analyse sémantique de texte par ordinateur, *Guidexpert-ato*¹, qui produit une émergence automatique des champs sémantiques et révèle ainsi les thèmes les plus populaires de discussion.

La présente analyse porte sur trois mille messages (cinq cent mille mots) échangés entre élèves et mentors dans la ressource de cybermentorat Academos². Academos met à la disposition de jeunes québécois des mentors provenant d'un grand nombre de métiers et de professions afin de les guider dans leur démarche vocationnelle. Les messages sont échangés par l'intermédiaire de la ressource web, ce qui assure leur archivage. Une évaluation approfondie de la satisfaction des élèves et des mentors a révélé que les participants apprécient la ressource. Il est cependant nécessaire de démontrer que les sujets de discussion portent vraiment sur des thèmes associés au cheminement professionnel de l'élève.

Guidexpert-ato génère automatiquement les champs sémantiques (liste de vocables ou de synapsies ayant des affinités sémantiques) et leur classement en ordre de fréquence. Des variables descriptives associées à chacun des messages permettent de comparer les champs sémantiques selon le sexe, l'âge, le statut d'élève ou de mentor, le nombre de messages, la durée de la relation, etc...

¹ Guidexpert-ato : <http://fable.ato.uqam.ca/guidexpert/guidexpert-ato-wp.htm>

² Academos : <http://www.academos.qc.ca>

L'analyse effectuée sur les messages révèle près de huit mille champs sémantiques. Les deux champs sémantiques les plus fréquents sont *le travail* et *le savoir*, présents dans plus de cinquante pour cent des messages. Des analyses différenciées selon le sexe ou le statut (mentor ou élève) retrouvent ces deux thèmes comme les plus fréquents dans chacun des sous-groupes. D'autres analyses révèlent la présence de thèmes associés plus particulièrement aux mentors (plaisir, chance) ou aux mentors et élèves féminins (vouloir).

La nature des champs sémantiques les plus fréquents confirme que les discussions ont bien porté sur des thèmes associés à l'orientation professionnelle. Ces résultats apportent une validation supplémentaire aux analyses de satisfaction des usagers. D'autres analyses devront être effectuées afin de désambigüiser les champs sémantiques en excluant des vocables non-pertinents et en regroupant certains champs sémantiques qui sont intimement liés.

L'analyse de thèmes de discussion ou de champs sémantiques par ordinateur semble particulièrement pertinente pour l'étude des courriels des communautés virtuelles de travail, de pratique ou d'entraide. La puissance de l'outil d'analyse sémantique provient de la présence d'un thésaurus des champs sémantiques des langues française et anglaise et ensuite de l'exécution rapide des requêtes. Des analyses qualitatives qui prennent plusieurs mois peuvent être ainsi effectuées en quelques minutes. De plus, comme les catégorisations émergent du logiciel plutôt que de la tête des chercheurs, l'analyse acquiert une plus grande objectivité.

Le commerce des écritures électroniques

Valérie Beaudouin

Chercheuse

France Télécom R&D
Laboratoire Usages, Créativité, Ergonomie
38-40, rue du Général Leclerc
92794 Issy-les-Moulineaux cedex 9

valerie.beaudouin@francetelecom.com

L'internet est un média qui articule de manière originale et inédite les propriétés des médias de masse (un producteur s'adresse à un public vaste et anonyme) et celles des outils de communication interpersonnelle (qui renvoient aux pratiques de sociabilité). Plus précisément, le réseau permet un entrelacement fin entre les pratiques d'écriture et de lecture.

En effet, l'écriture et par conséquent la lecture sont au cœur des pratiques de l'internet que ce soit pour la production et la réception de documents que pour la conversation. Il s'agit peut-être d'un état transitoire, mais c'est autour de l'écrit que s'est constitué le média. Un même média sert à des activités qui autrefois se répartissaient d'une part dans le domaine de l'écrit pour la publication (voir le monde du livre), d'autre part dans celui de l'oral pour la conversation.

Ce rassemblement inédit conduit à la coexistence de formes d'écriture extrêmement codifiées et diversifiées (livres, articles, échanges dans les forums, dans les chats...) et par conséquent à des formes d'appropriation qui varient sensiblement selon les compétences des acteurs (compétences linguistiques, techniques...) et leurs dotations en capital économique, culturel ou social.

Ce n'est pas tant dans l'accès ou non à internet que se construit le fossé numérique mais bien dans la manière de s'approprier et d'articuler les différents types d'activité sur le réseau. En partant de l'analyse des usages d'une cohorte d'internautes, nous montrons comment se construisent des types d'utilisation d'internet très distincts. On peut proposer un modèle étagé, où l'acquisition d'une compétence interactionnelle ne conduit pas à l'abandon des autres. Ainsi si la navigation sur le web est partagée par tous (avec un degré d'intensité très variable), l'utilisation régulière de la messagerie électronique est plus restreinte et les pratiques de conversation synchrones ne sont l'apanage que d'une petite frange d'internautes qui utilisent aussi beaucoup le web et le mail.

Nous montrerons ensuite que l'entrelacement entre la publication et la conversation, que l'on observe de manière très forte dans tous les collectifs qui se créent sur le réseau, est une condition de l'organisation territoriale d'internet : c'est à travers les échanges interpersonnels que se constituent la valeur et la hiérarchisation des documents qui y circulent.

Ces échanges, parce qu'ils sont sédimentés, mémorisés dans les espaces numériques, deviennent des ressources pour s'orienter dans la pluralité des sites. Avec le temps, les utilisateurs utilisent de plus en plus *l'Autre* comme ressource pour découvrir des documents plutôt que les services anonymes comme les moteurs de recherche ou les annuaires.

Internet, une cité imparfaite*

Richard Delmas

Administrateur principal à la Commission Européenne
DG société de l'information
Conseiller auprès du comité des gouvernements de l'ICANN

Richard.Delmas@cec.eu.int

Il est maintenant couramment admis que, du point de vue de la dynamique du changement des sociétés, internet joue un rôle emblématique sinon structurant. Le réseau des réseaux se situe résolument au croisement des sphères informationnelle, juridique et économique, à l'intersection des espaces privés et publics. On s'interroge cependant peu sur sa vraie nature : infrastructure globale, nouveau média, ou grand récit, raccourci commode de notre représentation du progrès technique ? Dès l'origine, le « réseau des réseaux » se présente sous une forme paradoxale : un univers en escalier, dont les paliers seraient suspendus à un axe sans véritable hiérarchie. Par construction néanmoins, la centralité et le contrôle du réseau sont implicites, au nom des principes d'unicité, d'intégrité et de sécurité, souci principal des agences fédérales américaines qui le gèrent. Mais internet est aussi fondamentalement distribué, interopérable, et donc vecteur d'autonomie pour ceux qui veulent s'affranchir des régimes de communication en place. On parle alors d'un espace unique de liberté, voire d'une utopie libertaire.

Internet se démultiplie dès le début des années 1990 avec la généralisation du *web*, c'est-à-dire des contenus en ligne, et de leur accès libre par moteur de recherche et par un système d'identifiant à vocation universelle, celui des noms de domaine. Cette moderne tour de Babel a engendré une spirale d'anticipation et de sidération suscitant une bulle marchande non maîtrisée. On a oublié que la présentation et la dénomination sur internet, c'est-à-dire le pouvoir de publier et de nommer sur la toile, comportaient une dimension sociétale, éthique et culturelle qui crée des droits et des contrats nouveaux. Aujourd'hui, le phénomène s'accroît, entraîné dans un processus d'actualisation apparemment sans fin, prêt à épuiser toutes les virtualités du protocole : nouvelle norme Ipv6 aux nombres quasi-infinis, *metadata*, *grid*, communications spatiales, etc. Mais en dépit de la profusion des lois et des textes portant sur l'un ou l'autre des aspects d'internet, on distingue mal le fondement du droit, la substance de règles qui seraient unanimement reconnues au plan international. internet demeure l'objet d'une pensée faible, orpheline, en quelque sorte.

* Les propos n'engagent que leur auteur et en aucun cas la Commission Européenne

Les groupes leaders des listes électroniques

Un essai de caractérisation

Madeleine Akrich et Cécile Méadel

Chercheuses

Centre de sociologie de l'innovation - École des mines de Paris
60 boulevard Saint-Michel
75006 Paris

akrich@ensmp.fr, meadel@ensmp.fr

Les cercles de discussion par messagerie électronique constituent un espace de débat radicalement nouveau à la fois dans son type d'échanges et dans les formes de liens entre participants. S'y installe en effet une parole au statut hybride, mi-publique puisqu'elle s'adresse à un groupe de gens qui ne se connaissent pas, mi-privée puisque la diffusion de ces messages est limitée à un groupe identifié de personnes. La participation à ces forums de discussion est d'une grande facilité : les obstacles sont faibles tant pour les aspects techniques que pour les procédures d'adhésion et les formes de contribution au débat. Or cette simplicité s'accompagne d'une implication forte. Ces caractéristiques en font des lieux d'échange riches et très originaux.

Les cercles de discussion électronique (CDE) sont donc particulièrement intéressants dans la mesure où ils proposent un espace nouveau de constitution, voire de reconfiguration, du lien social : établis autour d'un intérêt collectivement partagé pour un même objet, ces cercles donnent lieu à des échanges dont la variété des participants et des approches, la diversité thématique, la réactivité sont des caractéristiques majeures. Nous faisons l'hypothèse qu'au travers de ces listes de discussion se constituent actuellement de nouvelles formes d'expression de l'opinion publique, conduisant dans certains cas à l'émergence de véritables collectifs, distincts des formes traditionnelles comme les syndicats, les associations, les groupes de militants, les institutions professionnelles... La question de savoir le rôle que pourront jouer ces collectifs dans la cité reste ouverte, mais certains exemples montrent qu'ils ont pu déboucher sur des formes d'action collective.

Comment alors rendre compte des discussions qui se déroulent sur ces listes et des reconfigurations éventuelles qu'elles provoquent dans l'espace où elles s'inscrivent ? Les caractéristiques des CDE exigent des méthodes spécifiques d'investigation. En effet, les approches ordinaires en matière d'analyse du discours se heurtent ici à une double difficulté. D'une part, le matériau est constitué par des lettres de particuliers, mais celles-ci sont destinées à un ensemble d'individus ; ces messages s'inscrivent dans un débat, qui est nourri par un ensemble de contributions individuelles, tantôt successives et tantôt simultanées. Autant dire que le message existe à la fois dans son contenu spécifique mais aussi dans ce que nous appelons un fil de discussion, c'est à dire un ensemble de courriers qui se répondent et s'interpellent. D'autre part, obstacle pratique mais peut-être plus sérieux encore, l'ampleur des échanges rend illusoire l'espoir d'analyser ou de lire la totalité des messages d'une liste.

Aussi travaillons-nous à mettre au point un outil de mesure de la sociabilité des ces listes¹ à partir de thèmes liés à la santé et la médecine (et qui impliquent patients ou professionnels). Nous disposons ainsi d'un matériau de plus de 150 000 messages recueillis auprès d'une cinquantaine de listes francophones. Nous posons en particulier les questions suivantes : peut-on classer les listes en fonction du rythme des échanges et des types de participation ? Quelle sorte de « démocratie » se développe dans les listes (on connaît les nombreuses spéculations qui présentent le média internet comme un facteur de renforcement des pratiques démocratiques). Comment caractériser les liens que créent les listes entre les locuteurs ?²

Dans notre communication, nous proposons d'étudier le groupe des *leaders*, ceux-ci étant définis comme les locuteurs qui envoient un nombre élevé de messages à la liste. En effet, la répartition de la parole dans les listes est toujours très inégale. Comme en d'autres domaines (par exemple l'économie ou la science), on constate grossièrement que vingt pour cent des acteurs détiennent quatre-vingts pour cent de la richesse (si nous appelons richesse dans ce cas la prise de parole). Cependant, sur ce schéma globalement inégal, il existe différents modèles de répartition de la parole, qui dépendent du « format » de la liste (défini par le rythme et le volume des échanges ainsi que la distribution de la parole entre locuteurs).

L'analyse porte donc plus spécifiquement sur le groupe des « gros locuteurs », ceux dont le nombre d'interventions moyen est deux fois supérieur à celui de la moyenne des locuteurs de la liste (ce qui permet d'annuler l'effet des tailles contrastées des listes). Nous essaierons donc de définir quelques indicateurs qui permettent de caractériser leur activité et en même temps de dégager des « types » de listes qui seraient définies par certaines valeurs de ces indicateurs, étant entendu qu'il s'agit de prototypes ou d'idéaux-types. En particulier, on s'intéressera aux indicateurs suivants :

- le rapport entre la taille du groupe des gros locuteurs et l'activité de ce groupe ;
- la part relative des messages des gros locuteurs dans l'activité de la liste ;
- la part relative des messages de ce groupe dans les sujets collectifs (*i.e.* les sujets dont le nombre de locuteurs différents est supérieur à deux fois la moyenne de la liste).

Les mêmes indicateurs seront appliqués au *leader* (ou locuteur le plus actif) qui est susceptible de jouer des rôles différenciés selon les listes.

On confrontera cette caractérisation avec une caractérisation de l'activité de la liste dans son ensemble que nous avons décrite en particulier par deux autres indicateurs (la prolixité des locuteurs et le caractère collectif des échanges).

¹ Dans le cadre d'un projet de recherche CNRS-Plan du programme interdisciplinaire « Société de l'Information 2001 ».

² Petit point de vocabulaire : nous appelons « locuteurs » les personnes qui envoient des messages sur la liste : l'espace des locuteurs n'est pas tout à fait superposable à celui des abonnés dans la mesure où certains abonnés se contentent de recevoir des messages et n'en envoient jamais. On connaît éventuellement leur nombre par un message du modérateur de liste.

Evolution des métiers de la documentation

Alain d'Iribarne, Directeur de recherche (CNRS-LEST Aix en Provence)

Anne de Fenoyl, Conseil indépendante en RH

Marco Botta, Consultant Ergonomie (Stratexis)

Patrice Hemmel, Responsable Dvpt des Méthodes pour la gestion ressources humaines
(DRHC/DMEX Groupe TOTAL)

iribarne@univ-aix.fr, anne.defenoyl@ezagroup.com
marcobotta@compuserve.com, patrice.hemmel@totalfinaelf.com

Le niveau de diffusion des technologies de l'information et de la communication, dans les entreprises est aujourd'hui suffisant pour commencer à évaluer ce qui est appelé «l'impact» de ces dernières sur les transformations des métiers. C'est ainsi que Total a souhaité se doter d'une grille de lecture des changements en cours, en s'appuyant sur une analyse approfondie du métier de documentaliste. Ce choix est d'autant plus intéressant que d'une part, ce métier s'est considérablement complexifié par la multiplication et de la diversification des supports numériques et par la divulgation d'outils de recherche d'informations auprès de l'ensemble du personnel, et que d'autre part, ce métier est au cœur des processus de recomposition de la gestion de la connaissance.

C'est dans cette double perspective de transformation organisationnelle et technologique qu'ont été abordés les problèmes posés par l'usage de deux outils : un portail d'information largement diffusé auprès des *managers* et un moteur de recherche d'information sur l'internet réservé aux documentalistes. Trois axes majeurs de travail pour les gestionnaires ont pu être dégagés :

Un premier axe confirme l'hypothèse que les facteurs d'évolution de l'organisation se structurant en réseau ont pour effet d'exiger plus de transversalité dans les activités et donc de nouvelles formes de coopération. En terme de savoirs et de connaissances, si les « bases du métier » restent les mêmes on observe nettement un élargissement des bases professionnelles et des domaines d'interventions par le « déplacement du client » des documentalistes devenu lui-même, par acculturation aux outils, un utilisateur direct. Ainsi il ne leur suffit plus de rechercher ou de gérer l'information pour le compte du client, il leur faut aussi : former, auditer, analyser, faire de la veille pour accompagner ce dernier en partant de ses propres contraintes.

Un deuxième axe, autour de l'usage des outils, fait apparaître des niveaux de complexité des outils fort variables en fonction des modes de raisonnement requis, renvoyant aux formes de barrières d'entrée cognitives. Ainsi, le portail d'information, a-t-il pu être qualifié de «simple», car il mobilise un mode de raisonnement essentiellement déductif qui part de savoirs et de connaissances formelles et fonctionne par analogie sur la base de règles préétablies. A l'opposé le moteur de recherche sur l'internet a pu être qualifié de «complexe» car il mobilise des processus de raisonnement principalement inductifs, dans la mesure où il s'agit de «construire» la connaissance (thésaurus de l'utilisateur direct) en même temps que s'exerce l'activité : avec cet outil, les documentalistes sont conduits à passer de l'alimentation de bases de données structurées à la création de bases de connaissances mouvantes, au gré des changements de contexte et du profil de l'utilisateur. On mesure ici toute l'importance de l'apprentissage permettant aux personnes concernées de se construire leurs propres règles

d'usages. Dans cette perspective d'instabilité des connaissances, les documentalistes doivent mettre en œuvre des savoirs et savoir-faire bien supérieurs puisqu'il s'agit, en coordination étroite avec l'utilisateur, de rechercher l'information en définissant un profil unique d'espace sémantique simplifié, désambiguïté et enrichi avec les termes du métier de l'utilisateur direct.

Le troisième axe découle précisément des deux premiers et interroge l'identité du métier de documentaliste. D'un côté, le métier devient encore plus que précédemment un métier traversant c'est-à-dire susceptible d'être exercé par des personnes en provenance d'autres secteurs d'application, ce qui remet en cause les constructions de parcours professionnels incluant un passage dans la fonction. Inversement le métier devient aussi un métier «d'expert», pour des documentalistes capables de favoriser le déploiement des outils de plus en plus complexes. Se pose alors la question de la construction de filières d'expertise -dans le temps-, incluant l'élargissement des bases professionnelles avec les domaines d'intervention et la variété des techniques à maîtriser.

Dans ces dimensions, les critères de performance et la reconnaissance de celle-ci doivent eux-mêmes être réadaptés et réajustés aux modalités de travail des organisations en réseau. Ce sont donc tous les outils et mesures de gestion sociale qui sont à reprendre.

Études des Grilles et des Systèmes Pair-à-Pair :

de la mesure au grand instrument d'expérimentation

Olivier Richard

ENSIMAG

Olivier.Richard@imag.fr

Les recherches en informatique menées dans le domaine des grands systèmes distribués ont connu ces dernières années d'importantes évolutions. Ainsi la notion de grille de calcul (GRID) est apparue suite à la volonté d'utiliser conjointement les ressources de calcul de grande capacité (machines parallèles) réparties géographiquement. A l'échelle d'internet ont émergé les systèmes Pair-à-Pair dont l'objectif est la mise en commun d'un très grand nombre de ressources des machines de type PC, par exemple pour l'échange de fichiers (ex. Napster, Gnutella) et pour le calcul (ex. [Seti@home](#), Decryphon). Ces deux familles de système ont en commun le très grand nombre de ressources mises en jeu, ce qui rend difficile leur étude et l'élaboration rigoureuse de logiciels pour leur exploitation.

Ainsi, depuis les premières expériences du projet I-WAY aux États-Unis qui ont donné naissance au projet Globus, l'étude expérimentale des Grilles s'est essentiellement concentrée sur des déploiements à échelle réelle impliquant l'étude et le développement de logiciels système et d'intergiciels. L'approche orientée déploiement pose du point de vue scientifique plusieurs problèmes : a) la méthodologie suivie consiste à résoudre l'ensemble des problèmes simultanément (sécurité, recherche de ressource, allocation de ressource, communication entre les ressources, ordonnancement, stockage et transfert de données), b) chaque déploiement est unique et les résultats d'expérience sont difficilement généralisables. Enfin, le coût d'un déploiement est tel qu'il est souvent justifié par une perspective de mise en production et qu'il utilise comme nœuds de la Grille des machines déjà en production. Cette approche interdit pratiquement l'expérimentation informatique pour des problèmes de sécurité, de partage de ressources, de conditions expérimentales et de reproductibilité des résultats.

Les études des systèmes de Pair-à-Pair sont tout aussi délicates : la maîtrise d'un déploiement est illusoire au-delà d'une centaine de nœuds et le problème de la reproductibilité des résultats est accentuée par une très forte variabilité des conditions d'expérimentation (trafic internet, disponibilité des machines participantes).

Ces difficultés sont à l'origine d'une initiative nationale (projet Grid5000) qui a pour objectif la mise en place d'une plate-forme dédiée à l'expérimentation et à l'étude des Grilles et des systèmes Pair-à-Pair.

Le projet vise à proposer une réponse à ces problèmes en spécifiant les caractéristiques et les modalités de fonctionnement/utilisation d'une Grille purement expérimentale connectant une dizaine de sites géographiquement distribués et reliés par des réseaux à haut débit. Il faut percevoir une Grille expérimentale comme un «grand instrument» à destination des informaticiens. A terme, une telle plate-forme devrait recevoir une charge expérimentale pour

l'étude de mécanismes de Grille et de systèmes Pair-à-Pair. Il s'agirait aussi de tester des systèmes d'exploitation, des intergiciels, des applications, des techniques d'observation/maintenance.

Parmi les différents efforts entrepris dans ce projet, la constitution d'une base de mesures est un élément central. Elle consiste à rassembler des mesures caractérisant l'environnement des Grilles et des systèmes Pair-à-Pair. A titre d'exemple il a été initié le recueil et l'analyse des mesures de débits observées par les internautes utilisant l'ADSL via le logiciel «*La grenouille*». Ces mesures permettent d'avoir une cartographie riche d'information sur l'ensemble du territoire que forme ce réseau d'utilisateurs. D'autres mesures, comme le recueil des requêtes traitées par les systèmes Pair-à-Pair pour les échanges de fichiers, permettent de caractériser le comportement de leurs utilisateurs.

Une exploitation possible pour ces mesures est leur utilisation comme *stimulus* des systèmes à évaluer. La méthodologie qui est envisagée est celle de l'émulation. Elle consiste à reproduire dans un environnement complètement maîtrisé les conditions d'expérimentations réelles et d'exécuter en son sein le système à étudier. Cette piste permet d'envisager des études avec un niveau de réalisme élevé qui seront complémentaires de celles fondées sur des approches classiques par modélisation et simulation. Mais surtout elle devrait répondre à l'impossibilité actuelle d'entreprendre des tests par déploiement réel de grande taille.

Mesurer les usages de P2P ?

Geneviève Vidal, Maître de Conférence
LabSic, Université Paris 13, MSH Paris Nord

Vincent Mabillot
Médias et Identités, Université Lyon II, IUT Grenoble

gvidal@sic.univ-paris13.fr, vincent.mabillot@free.fr

Au gré des crises, techniques et juridiques, qui secouent les réseaux de *peer-to-peer* (P2P) sur internet, nous constatons des déplacements d'usagers vers de nouvelles offres de services. En effet, les internautes contournent les obstacles en s'appuyant sur leurs compétences informatiques. Cette mobilité exige des tactiques, voire des stratégies de transmission d'information, de savoir-faire, et l'acquisition de compétences technique et communicationnelle.

Pour comprendre les conduites de partage des informations et les pratiques, une double question se pose : qui sont les usagers de ces espaces d'échanges soumis à des évolutions rapides ? Comment passent-ils d'une technologie à une autre, installent-ils et découvrent-ils de nouveaux logiciels, de nouveaux modes d'usages ?

Suite à la diffusion en ligne d'un questionnaire, durant deux mois (octobre et novembre 2002), nous avons recueilli 575 réponses, dont 359 ont été traitées statistiquement. Leur analyse permet de mettre en lumière des pratiques collaboratives, affirmant des modèles relationnels sans finalité mercantile entre les acteurs. Tout en restant mobiles pour maintenir la pratique dans un contexte socio-économique instable, les internautes font émerger des normes d'usages.

Nous avons établi des profils d'utilisateurs, des modes d'acquisition de compétences, des stratégies d'apprentissage et de mobilité. En premier lieu la diversité des démarches relatives aux stratégies d'apprentissages techniques nous a conduit à établir un parcours d'apprentissage moyen : télécharger, installer, lancer le logiciel et l'utiliser de façon inductive. Lorsque les utilisateurs rencontrent des difficultés d'utilisation, ils préfèrent consulter internet plutôt qu'un mode d'emploi pour trouver des explications. En cas d'insuccès, ils cherchent par le biais des messageries une personne ressource susceptible de leur donner des conseils avisés. De façon plus exceptionnelle, ils consultent des ressources hors internet (livres, magazines). En tout dernier lieu, les utilisateurs renoncent, mais nous supposons alors qu'ils changent de logiciel de P2P.

Mais ce profil moyen est assez peu satisfaisant pour rendre compte de la diversité des démarches. En effet, on a mis en évidence une logique d'apprentissage hésitante et timide pour ceux qui n'ont qu'un logiciel, une approche concentrée sur internet et demandeuse d'assistance pour ceux qui découvrent les potentialités du réseau et enfin une approche distanciée pour les « experts ». Nous pouvons en outre saisir le fait que les novices essaient d'acquérir un minimum de connaissances pour pouvoir formuler leur demande. Les nouveaux adeptes (2-3 logiciels utilisés) essaient d'acquérir un statut de compétence et de reconnaissance dans les groupes d'utilisateurs. Quant aux experts, ils ont franchi toutes les

étapes d'apprentissage et n'utilisent plus les systèmes de messagerie pour apprendre, mais pour communiquer autour du P2P (pour négocier, découvrir des contenus, discuter).

Enfin, l'analyse des stratégies de transmission d'information, d'acquisition de compétences technologiques et communicationnelles pour rester mobile en cas de crise (définie comme des ruptures et des transferts d'usages), atténue la dimension d'un réseau d'individus qui coopèrent pour se déplacer vers d'autres dispositifs.

Nous concluons provisoirement sur le fait que l'expérience de crise amène les utilisateurs à développer des capacités d'adaptation à d'autres dispositifs et augmente ainsi le niveau d'expertise et le recours à la multiplication des solutions de replis.

La poursuite de cette étude devrait permettre une meilleure compréhension des échanges entre les individus, de leurs rapports à l'apprentissage et de la capacité des groupes d'internautes à s'organiser sur internet.

« Mesurer » des situations de sociabilités médiatisées

Paulette DUARTE

Sociologue-urbaniste, M.de Conf. à l'Institut d'Urbanisme, d'Aménagement et d'Administration de Grenoble (Univ. Pierre Mendès France, Grenoble), Chercheur au lab. Territoires de l'UMR PACT (Politiques Publiques Actions Politiques et Territoires)

Mohammed BOUBEZARI

Architecte chercheur (Instituto Superior Técnico, Universidade Técnica de Lisboa)

Marie-Christine COUIC

Sociologue-urbaniste, Activité libérale, collectif BazarUrbain, Enseignant et chercheur contractuel à l'Institut d'Urbanisme, d'Aménagement et d'Administration de Grenoble de l'Université Pierre Mendès France de Grenoble

paulette.duarte@upmf-grenoble.fr, m.boubezari@ist.utl.pt, mc.couic@wanadoo.fr

Cet article présente une méthode d'observation et d'analyse des sociabilités médiatisées, basée sur la notion de situation, comme exemple de ce qui est mesurable *via* internet, ainsi que des résultats d'observation et d'analyse. Cette méthode utilisée dans la recherche « Le Médiat et l'Immédiat dans les espaces de sociabilité contemporains¹ » est exploratoire, microscopique et qualitative. Elle vise à « mesurer » le sens de la sociabilité médiatisée et les types de sociabilité contemporains par ce que nous donne à voir l'internaute.

Pour étudier les sociabilités médiatisées par internet, nous avons choisi de prendre du recul par rapport aux échelles traditionnelles (logement, quartier, ville, agglomération...) et d'observer les liens entre des espaces publics et des espaces privés. Pour nous, les situations de sociabilités médiatisées par internet sont des situations urbaines, elles en ont les mêmes caractéristiques et elles peuvent être médiates ou immédiates et contribuer à rendre plus poreuses les frontières entre sociabilités médiatisées et domestiques.

Cette recherche vise, non pas à « mesurer » numériquement les sociabilités médiatisées à partir de l'espace domestique, mais plutôt 1) à évaluer de manière qualitative et exploratoire la forme et la nature des sociabilités actuelles qui ont lieu sur internet (celles qui se créent, s'actualisent ou se renforcent *via* l'internet, grâce aux e-mails, « chats », etc.), 2) à les typer, 3) mais également à étudier, dans une visée prospective, la transformation réciproque du logement (agencements, modes de relations familiales...) et de l'espace public (lieux de rendez-vous, usages...).

La méthode, dans la phase d'observation, concilie différentes techniques d'enquête complémentaires : entretiens semi-directifs avec des internautes qui, de par leurs pratiques,

¹ Cette recherche est inscrite au programme de recherche du Plan Urbanisme Construction et Architecture du ministère de l'Équipement, du Logement et des Transports : « Habitat et Vie Urbaine ». Ce programme souhaite faire, entre autres, le point sur les espaces publics urbains, les espaces domestiques et la pertinence des échelles actuellement utilisées pour les décrire.

possèdent un savoir singulier sur leur sociabilité, et enregistrements vidéographiques sur leur lieu de pratiques ; observation directe et indirecte *in situ* de dispositifs qui favorisent la rencontre spatialement ancrée (LAN party) ; et enfin observation directe participante d'espaces de sociabilité de fait où les internautes se connectent avec l'intention de rencontrer d'autres internautes (Salons, « Rooms » dans un IRC). Ces techniques, éprouvées par les sciences sociales, sont soit mises en œuvre en même temps, ou l'une après l'autre, dans une situation de sociabilité vécue, soit mises en œuvre dans des temps et des lieux différents pour reconstituer cette situation de sociabilité vécue. Notre méthode est centrée sur l'internaute. C'est par ce dernier que nous avons accès aux données. Il nous les énonce ou nous les montre.

Cette méthode permet de croiser des matériaux de nature différente (matériaux discursif, graphique et vidéographique), de sélectionner les données pertinentes issues de ces matériaux en les articulant, les juxtaposant, voire en les superposant en partie. Les résultats obtenus se présentent sous forme de fiches de situations-types de sociabilité dégagées : situations-type famille, rencontre, contacts sensibles, communautaire, partage, visite, « mailer », « mots-dits », immédiateté, ubiquité-composite. On comprend alors que la sociabilité médiatisée a pour une grande part les mêmes caractéristiques que la sociabilité classique contemporaine : contradictoire, complexe, éclatée, diffuse. Contradictoire et complexe, elle est, par exemple, à la fois médiante et immédiate, publique et privée. Eclatée et diffuse, elle l'est car, par exemple, l'internaute entretient des relations sociales avec des internautes qui sont distants ou proches géographiquement mais qui ne se connaissent pas entre eux, et parce que l'internaute entretient des relations à la fois avec des internautes et des membres de son espace domestique.

Ces situations-types de sociabilité, bien que pour la plupart ancrées dans l'espace domestique, montrent que la qualité de l'espace varie dans le temps, devenant public, privé, ouvert, fermé... et que cet espace vécu composite est celui de la sociabilité de l'internaute. Autrement dit, ces qualités de « public ou privé, ouvert ou fermé, etc. » ne sont pas des attributs exclusifs de l'espace construit mais appartiennent à la situation avec ce qu'elle comporte de perceptions d'actions et de représentations en plus de l'environnement physique. L'espace est à la fois physique, perçu et représenté. C'est un espace sensible, expérimenté et connu par l'internaute même quand il est médiatisé, un territoire.

Par ailleurs, l'aspect temporel des situations-types nous révèle des habitudes, des expériences renouvelées. Les situations mettent en évidence des aspects accélérant ou ralentissant, obstruant ou exposant, dilatant ou rétrécissant des expériences sensibles des individus et dont la seule observation de l'espace ne saurait en rendre compte.

Au final, et à la lumière des résultats obtenus, il nous est donné de constater que la sociabilité se re-construit et se reproduit à travers le média internet malgré les obstacles ou les performances de ce dernier ; car pour se faire, elle se base sur la compétence ordinaire de l'internaute. Sous cet angle, l'internet n'est pas un objet, mais un lien social.

Invariance d'Échelle dans l'Internet

Patrice Abry et **Patrick Flandrin**
Laboratoire de Physique, CNRS-ENS Lyon, France

Nicolas Hohn et Darryl Veitch
Dept. of EEE, University of Melbourne, Australia

www.ens-lyon/~pabry
www.ens-lyon/~flandrin
www.emulab.ee.mu.oz.au/~darryl

Signaux — Le réseau internet se caractérise par une très forte hétérogénéité (géographique, topologique et matérielle) ainsi que par une importante diversité des flux d'informations (données, applications) qui y circulent. Ces flux sont eux-mêmes l'objet de nombreux conditionnements (fragmentations en paquets IP, multiplexages, routages), faisant de la mesure du trafic en un point du réseau une source de signaux d'une grande complexité. Mieux connaître le comportement statistique de ces signaux et être capable de les modéliser, voire d'en proposer des mécanismes générateurs, apparaît ainsi comme un point de passage obligé pour une meilleure maîtrise et un meilleur contrôle de l'activité du réseau en termes de dimensionnements, de qualité de service ou d'élaboration de nouveaux protocoles.

Invariance d'échelle — L'ensemble des activités présentes sur internet couvre un large éventail d'échelles de temps, allant de la milli-seconde (ordre de grandeur lié à la technologie et aux protocoles) à l'heure (activité humaine et usages). Depuis les premières mesures effectuées à la fin des années 80, il est apparu qu'aucune de ces échelles de temps ne jouait de rôle privilégié ou, en d'autres termes, que le trafic internet était invariant d'échelle. Du point de vue de la mesure (ou de l'analyse du signal), cette situation pose deux questions essentielles. La première est de disposer d'outils adéquats pour mettre en évidence et quantifier l'invariance d'échelle, la deuxième étant de pouvoir proposer des modèles générateurs.

Ondelettes — L'invariance d'échelle se traduisant sur les signaux par une auto-similarité statistique (fractale aléatoire), il est naturel de se tourner vers des outils susceptibles de scruter un signal à différentes échelles d'observation. C'est précisément ce que permet l'analyse en ondelettes, autour de laquelle ont été développées des méthodes performantes et statistiquement fiables pour l'estimation du (ou des) paramètre(s) caractérisant l'invariance d'échelle.

Longue mémoire — L'invariance d'échelle présente dans le trafic internet double l'auto-similarité statistique d'une propriété de longue mémoire, indiquant que des liens statistiques significatifs se maintiennent entre des fluctuations de trafic même très éloignées dans le temps. Ce fait a un impact considérable en termes de performances (retards), de qualité de service (perte de paquets) et de coûts (dimensionnement des systèmes).

Modèles — La présence essentielle d'auto-similarité et de longue mémoire rend inadaptés les modèles classiques (de type markovien) utilisés pour les systèmes de communication traditionnels comme le téléphone.

Des modèles «boîte noire» (par exemple de type brownien fractionnaire) existent alors, mais sans lien nécessaire avec des mécanismes générateurs du trafic. Un premier modèle effectif permet cependant de rendre compte de monofractalité aux grandes échelles sur la base d'un trafic *on-off* agrégé dont les distributions des durées d'activation et d'inactivation sont «à queues lourdes», renforçant ainsi la probabilité d'événements extrêmes.

D'autres mécanismes restent néanmoins à élucider pour rendre compte des caractéristiques plus riches observées sur des traces réelles - caractère non gaussien, (multi-) fractalité à petite échelle - et pour incorporer des contraintes liées aux protocoles et aux usages.

Systèmes complexes — L'internet est un exemple de «grand système complexe» lié à l'activité humaine, dont les caractéristiques d'invariance d'échelle se rapprochent néanmoins de celles observées dans d'autres systèmes «naturels» comme en turbulence, en géophysique ou en biologie. Le caractère fortuit ou profond de ce rapprochement est une question ouverte.

En savoir plus

Références — Pour une présentation générale relative aux lois d'échelle dans le télétrafic informatique, on pourra se reporter à D. Veitch, "Lois d'échelle en télétrafic informatique", Chapitre 6 du Volume 2 de Lois d'échelle, Fractales et Ondelettes (P. Abry, P. Gonçalvès et J. Lévy Véhel, eds.), Traité IC2, Hermes, Paris, 2002.

Une présentation tutorielle plus technique et plus précise de l'analyse des lois d'échelle par les ondelettes est donnée dans P. Abry, R. Baraniuk, P. Flandrin, R. Riedi and D. Veitch, "The multiscale nature of network traffic: Discovery, analysis, and modelling," IEEE Signal Processing Magazine, Special Issue on "Analysis and Modeling of High-Speed Data Network Traffic", vol. 19, no. 3, pp. 28--46, May 2002.

P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data," Chapitre 2 de Self-Similar Network Traffic and Performance Evaluation (K. Park and W. Willinger, eds), Wiley (Interscience Division), 2000.

Liens — Une bibliographie plus complète et des procédures Matlab d'analyses des lois d'échelle dans des données expérimentales sont disponibles sur les pages web des auteurs.

Topologie d'internet et du web : mesure et modélisation

Jean-Loup Guillaume et **Matthieu Latapy**

LIAFA - Université Paris 7
2 place Jussieu
75005 Paris

(guillaume,latapy)@liafa.jussieu.fr

L'interconnexion des machines sur internet et les liens hypertextes entre les pages web structurent internet et le web.

Ces structures peuvent se modéliser par des graphes, qu'il s'agit d'une part d'explorer (mesure) et d'autre part de comprendre et de reproduire (modélisation).

Nous discutons dans cette communication les problématiques soulevées par l'exploration de ces graphes ainsi que les premières propriétés qu'on leur connaît.

Nous terminons par une présentation des principales propositions faites pour les modéliser.

L'objectif, plus qu'un exposé technique des faits, est de donner une vue d'ensemble de la thématique.

Écriture, informatique et internet

Réal Gingras

Prof en ligne

apel@cam.org

<http://profenligne.cam.org>

À partir des protocoles internet utilisés chez *Prof en ligne* depuis 1996, une relation étroite s'est établie avec les élèves; d'une part avec le courrier électronique, les activités proposées et les différents forums et d'autre part avec les contacts directs par « chat » et par vidéoconférence. Le prix de la francophonie canadienne reçu en 1998 est venu confirmer que l'approche numérique proposée était une valeur ajoutée à ce que nous pouvons appeler maintenant l'intégration pédagogique des TIC ; autant en classe qu'en ligne.

Fondamentalement, la présence en ligne assure une régularité dans les réponses à donner et dans les activités à proposer ou à recevoir. C'est cette régularité de la présence qui assure la qualité de la relation entre Prof en ligne et les élèves. Depuis 1999, nous avons, à partir de l'école Félix-Leclerc¹, commencé à installer des serveurs d'application qui roulent sous Linux et qui nous permettent de faire fonctionner des ordinateurs considérés comme désuets.

La grande stabilité et l'universalité de ce système nous permet de réaliser encore plus toutes les fonctions dont nous avons besoin dans le développement des TIC et des différentes interventions de *Prof en ligne*.

L'entraide et le partenariat sont au cœur de la relation et la connaissance devient distribuée. Linux offre donc, de par son ouverture, une liberté dans la création de contenu car il enlève toute contrainte que les systèmes marchands tentent de nous imposer.

Prof en ligne et Linux se rejoignent dans une même conquête: celle des réseaux libres nécessaires à l'apprentissage de la citoyenneté, si important dans la compréhension de notre monde de plus en plus numérique.

Autrement dit, nous ne sommes plus dépendants des réseaux propriétaires et toutes les latitudes créatives sont autorisées pour mettre en place des compétences à la fois stables et évolutives.

Souhaitons que la structure scolaire sera en mesure d'y voir là des outils qui ne demandent qu'à se développer...

¹ <http://fleclerc.csdm.qc.ca>

Ménestrel, un réseau mesuré

Christine Ducourtieux,

Ingénieur d'études

Laboratoire de Médiévisitologie Occidentale de Paris - Université Paris I

Christine.Ducourtieux@univ-paris1.fr

Ménestrel est un réseau mesuré, à la croissance contrôlée, qui n'a cependant pas boudé quelques grandes idées ravivées par l'arrivée d'internet dans le monde des historiens. Ménestrel est un réseau de médiévistes et de professionnels de l'information, né du constat qu'un nouveau support documentaire étant apparu, tout chercheur ou documentaliste sérieux ne pouvait l'ignorer et que, plutôt que de subir les changements induits par l'innovation technologique, il fallait accompagner et expérimenter cette dernière.

Et pourtant dans un premier temps, il y a la prise de conscience rapide de la différence de taille des deux protagonistes en présence : un réseau planétaire est imposé à une communauté accoutumée à borner ses investigations au monde d'avant les Grandes Découvertes ; un univers de grands chiffres, de superlatifs rencontre une communauté d'hommes accoutumés à un vocabulaire précis, voire austère ; le premier loue la vitesse, l'exhaustivité, caresse quelque désir d'universalité ; l'autre aime, la lenteur, la rareté et la richesse de l'exception.

Dans un second temps, internet, par le truchement des moteurs de recherche qui donnent à voir ses ressources, apparaît aux yeux du néophyte, à l'image d'un monde plat, sans hiérarchie avouée. Il crée une familiarité de traitement inhabituelle entre l'amateur et le spécialiste, générant l'illusion d'un monde égalitaire. A considérer la persistance et la vitalité des modèles académiques, cette illusion peut être funeste pour l'étudiant internaute enthousiaste qui espère ainsi approcher des chercheurs !

Ménestrel, avant de mesurer l'influence d'internet dans le monde des historiens ou de mesurer sa propre notoriété, a choisi la voie de la prudence et de l'expérimentation. Une croissance réussie, au regard de sa longévité et de sa prospérité ; cinq années déjà et bien des métamorphoses rendues nécessaires par une meilleure connaissance du formidable support d'expérimentation que constitue le réseau, par ce frottement obligé à d'autres disciplines et par l'adaptation à la multiplication des entreprises sur internet. Un réseau, tellement occupé à se construire, qu'il ne s'est pas soucié de se mesurer.

Ce qui explique cette méfiance à l'égard des bilans chiffrés mais ne suffit pas à expliquer cette réticence, voire cet oubli que des observateurs pourraient, non sans fondement, qualifier de volontaire. Les statistiques proposées nous apparaissent singulièrement étrangères à notre pratique et souvent rudimentaires : le nombre de visiteurs, les pages consultées, etc. Autant d'outils de mesure extrêmement grossiers qui ne décrivent que fort mal les usages développés avec une réelle inventivité par des internautes.

Nous avons conscience que nous ne pouvons pas continuer à décrire une expérience sans fournir à l'appui des preuves. Cet effort est indispensable si nous voulons que les travaux

réalisés, sur internet, soient validés scientifiquement et permettent de plaider, auprès des décideurs, la création de nouveaux métiers. Toutefois, nous espérons des instruments affinés qui puissent saisir l'activité d'un réseau dans sa dimension sociologique. Comment mesurer un réseau qui, par définition, privilégie le dynamisme des échanges et des entreprises et ne songe pas à thésauriser ?

Une non-mesure de l'internet : usages subjectifs et rêvés d'une bibliothèque numérique, contes et mécomptes d'un bibliothécaire

Olivier Bogros

Médiathèque André Malraux
Lisieux

bib_lisieux@compuserve.com

Liminaire

La Bibliothèque électronique de Lisieux est un service en ligne limité aujourd'hui encore, sept ans après son ouverture, à un simple réservoir de textes littéraires et documentaires du domaine public francophone alimenté à partir des collections anciennes conservées à la Médiathèque André Malraux de Lisieux. Ce service, qui peut être considéré comme une animation littéraire du web, a quelque prétention à faire des propositions de lecture et n'oublie pas que les bibliothèques, même sur l'internet, restent des pourvoyeuses de textes. Ce service, enfin, souhaite prendre en compte une donnée nouvelle, que chaque bibliothécaire peut constater jour après jour dans son établissement : à savoir que commencent à être de plus en plus nombreux ceux qui utilisent le web, à tort ou à raison, comme le lieu quasi unique de leur recherche d'informations.

Pourquoi construire des bibliothèques électroniques textuelles ?

La lecture tout simplement, la lecture plaisir et curieuse qui naît de la découverte de textes oubliés et rares. Nous n'évoquerons pas ici les débats récurrents sur l'impossibilité de lire sur un écran. L'internaute fait comme il veut, les textes sont à sa disposition. Pour des usagers empêchés par le handicap physique, la distance ou plus simplement par les horaires d'ouverture des bibliothèques publiques, les collections textuelles en ligne sont des substituts essentiels aux exemplaires imprimés inaccessibles conservés par les bibliothèques. Il en est de même pour l'enseignement du français langue étrangère qui fait de plus en plus appel aux ressources disponibles sur le réseau : cours en ligne, dictionnaires, outils de correction, réservoirs de textes numériques... Enfin les bibliothèques numériques sont ouvertes au travail scientifique. Des logiciels spécialisés permettent des recherches sophistiquées dans des corpus numérisés en mode texte.

Le constat des statistiques:

La Bibliothèque électronique de Lisieux est aujourd'hui bien référencée et en profondeur dans les principaux moteurs de recherche et sur les pages de liens de nombreux autres sites, notamment universitaires. La fréquentation du site est maintenant stabilisée et d'un mois sur l'autre sensiblement identique. Si la lecture des statistiques, ici pour mars 2003, d'accès au site (*access log*) n'est pas de nature à nous renseigner directement sur les usages

réels par l'internaute des textes mis à sa disposition, elle nous instruit par contre sur son comportement en ligne et surtout nous amène à relativiser fortement l'impact supposé du site.

Le rapport de statistiques identifie 108 pays d'ou proviennent les connexions de la France (19.332) à la Zambie (1). Assez logiquement, les connexions proviennent majoritairement du territoire français et des pays francophones. Si l'on s'en tient uniquement aux chiffres bruts de consultation (59.050 visiteurs et 120.094 pages vues) on ne peut être qu'impressionné, à notre échelle, par la fréquentation du site, bien qu'il faille déjà minorer ces résultats en prenant en compte l'action des moteurs de recherches. L'indication du nombre de pages affichées par visiteurs (2,03 en moyenne) conjuguée à la durée du passage sur le site (189 secondes en moyenne) ramène à une réalité plus tangible et montre bien l'effet «zapping» du web. On peut donc raisonnablement penser que le nombre des visites «pertinentes» sur le site est plus proche de 14.000 que de 59.050. Plus de 450 pages-textes différentes sont consultées tous les mois de 50 à plus de 8.000 fois. Tout naturellement c'est la page d'accueil du site qui arrive en tête suivie assez logiquement aussi de la page sommaire qui liste les auteurs et les œuvres disponibles sur le site. Ensuite et sans surprise, parmi les 20 premières, on retrouve les pages index de chaque rayon de la bibliothèque numérique (littérature, documentaires, archives, Normandie) et les pages consacrées aux auteurs les plus connus : Maupassant, Zola, Daudet, Baudelaire, Flaubert... ainsi que la sélection mensuelle (en mars 2003 un conte badin de Catulle Mendès). Un suivi précis de mois en mois de cet indicateur permettrait de se rendre compte si ce classement des meilleures pages consultées se renouvelle sensiblement ou au contraire se répète mensuellement. Une étude approfondie des statistiques donnerait aussi des indications précieuses sur les pics de fréquentation liés à la mise en ligne des nouveautés et leur annonce sur les listes de diffusion. Elle indiquerait comment parfois un réservoir de textes patrimoniaux peut coller à l'actualité. Enfin, plus de la moitié des connexions sur le site se faisant à partir des moteurs de recherche, il serait intéressant de mesurer le classement des pages du site en fonction des mots clefs utilisés dans les requêtes et de déterminer le nombre de visites à classer comme des erreurs de cliquage.

On devine donc que les usages des bibliothèques numériques peuvent être multiples : de l'amateur curieux se promenant dans les rayons virtuels de la bibliothèque à l'internaute pressé qui procède via les moteurs à une exploitation ciblée de l'information.

Usages des bibliothèques numériques en ligne

Houssem Assadi et **Thomas Beauvisage**

France Télécom R&D - DIH/UCE
38 rue du Général Leclerc
92794 Issy Les Moulineaux, France

{houssem.assadi, [thomas.beauvisage](mailto:thomas.beauvisage@francetelecom.com)}@francetelecom.com

Objectifs :

Le projet BibUsages, mené en 2002 par France Télécom R&D en partenariat avec la Bibliothèque Nationale de France, a pour objectif l'étude des usages de la bibliothèque électronique en ligne Gallica. De tels usages, innovants, s'insèrent aussi dans des pratiques stabilisées (au sein de la population des enseignants et des chercheurs, mais également auprès du grand public). L'accès immédiat à un corpus volumineux permet à des chercheurs d'envisager des études inédites, car techniquement impossibles auparavant. Par ailleurs, les enseignants (du collège au premier cycle universitaire) trouvent dans les bibliothèques électroniques une ressource inestimable pour l'élaboration de leurs cours.

Notre propos est d'une part d'expliquer, par des méthodes issues des sciences sociales et cognitives, des usages déjà largement diffusés (il existe déjà plusieurs bibliothèques électroniques librement consultables sur le web), mais dont une compréhension plus rigoureuse permettrait d'élargir la palette des fonctionnalités proposées (en les adaptant aux besoins et aux caractéristiques des utilisateurs). D'autre part, il s'agit de développer une approche «centrée utilisateur», rarement mise en œuvre dans les études d'usages sur le web.

Organisation du projet

Le projet, d'une durée de 12 mois, s'est déroulé en trois étapes :

1). Enquête en ligne sur le site de Gallica (mars 2002).

Un questionnaire a été soumis aux visiteurs du site Gallica en mars 2002 durant trois semaines. Il a permis d'avoir une connaissance plus précise du public de Gallica, et de recruter des volontaires pour faire partie du panel d'utilisateurs dont le trafic web a été enregistré.

Outre les caractéristiques socio-démographiques des répondants, le questionnaire s'articule autour de deux thématiques principales :

D'une part, l'usage de Gallica (fréquence des visites, rubriques consultées, etc.), et d'autre part les usages d'internet en général (intensité d'usage, services utilisés, types de sites visités, etc.).

Au terme de cette première étape, 2340 personnes ont répondu au questionnaire, et 589 ont accepté de faire partie du panel d'utilisateurs.

2). Constitution d'un panel d'utilisateur, installation du dispositif de capture de trafic chez les utilisateurs du panel et recueil des données.

Au terme de la procédure d'inscription et d'installation, le panel est composé de 72 volontaires dont les caractéristiques socio-démographiques

correspondent à celles de l'ensemble des répondants à l'enquête. Les données d'usage de ce panel ont été rapatriées sur un serveur centralisé de traitement de juillet à décembre 2002.

3). Conduite d'entretiens avec un échantillon d'utilisateurs volontaires faisant partie du panel (octobre 2002).

Ces entretiens ont concerné 16 des 72 participants du panel, et ont été axés autour de trois problématiques particulières : leurs usages d'internet en général, leurs usages des bibliothèques numériques et de Gallica, et les liens avec les pratiques de lecture et culturelles «off-line».

L'analyse croisée des trois sources de données - questionnaire en ligne, données de trafic, entretiens - permet ainsi de dresser un panorama des usages riche et dépassant les pratiques on-line proprement dites.

Résultats du projet

En premier lieu, cette recherche a permis de mieux appréhender les utilisateurs des bibliothèques électroniques. Elles attirent un public qui n'est pas forcément habitué aux bibliothèques, mais qui y vient par le biais de recherches spécifiques : dans les entretiens autant que dans le trafic observé chez les participants de l'étude, les fonds numérisés permettent de disposer de manière simple et rapide de documents de référence, difficilement trouvables, qui s'inscrivent dans le cadre de contextes de recherche précis. Ce public semble assez différent de celui des bibliothèques classiques, et les chercheurs «professionnels» y sont comparativement peu représentés. Les plus de quarante ans, actifs ou retraités, sont majoritaires dans la population observée, et les bibliothèques électroniques sont avant tout pour eux utilisées dans le cadre de recherches personnelles. L'intensité d'usage est ici bien supérieure à celle de la population générale des internautes français, et va de pair avec un très fort taux d'équipement en haut débit (câble, ADSL).

Ensuite, elle a permis d'appréhender les contextes d'usages des fonds numérisés : si, d'une manière générale, les utilisateurs des bibliothèques électroniques sont également de forts consommateurs de «contenus à lire» (journaux en ligne en particulier), au sein des sessions, l'usage des bibliothèques numériques est fortement corrélé à celui des moteurs de recherche d'une part, des sites de vente de biens culturels d'autre part. Deux profils se dégagent : celui du «chercheur amateur», dont les centres d'intérêt sont pointus et déjà bien connus de lui, et celui du bibliophile, pour qui Gallica fait office de catalogue avant achat. Dans les deux cas, la lecture en ligne est rare, tout autant que l'impression des documents téléchargés, et la lecture s'apparente à la recherche de fragments ciblés au sein de vastes collections laissant de côté la totalité des œuvres. Dans ce cadre, le statut des documents en ligne semble remis en cause : tandis que l'édition papier reste du côté de l'œuvre, l'édition électronique s'apparente à l'usuel.

Enfin, l'étude montre les passerelles entre web marchand et non marchand pour les utilisateurs. Alors que les acteurs d'internet (fournisseurs de contenus et d'accès) perçoivent une dichotomie forte entre ces mondes, les internautes passent indifféremment d'un type de site à un autre, et l'on doit plutôt parler d'enrichissement mutuel entre marchands et non marchands dès lors qu'on les envisage sous l'angle des pratiques.

Dès lors, attirant de nouveaux publics, induisant de nouveaux modes d'appréhension des textes, s'inscrivant dans des parcours de lecture inédits, les bibliothèques électroniques, loin d'être une simple version numérisée des fonds, s'apparentent à un nouvel espace de lecture et de consultation aux côtés des bibliothèques traditionnelles.

La cartographie d'internet, outil de mesure et d'enjeux

des dynamiques de territoire

Sophie Houzet et Loïc Grasland

UMR 6012 – ESPACE, Avignon
74, rue Pasteur
84029 Avignon Cedex 01

sophie.houzet@voila.fr, loic.grasland@univ-avignon.fr

La cartographie constitue un bon outil d'exploration et de mesure des enjeux de territoires de l'internet. Ces territoires recouvrent autant des configurations physiques de réseaux dans l'espace géographique que des services aux entreprises - dont l'activité est en totalité ou en partie liée à ces réseaux - ou encore des usages grand public. Un des intérêts particuliers de cette cartographie réside dans la possibilité d'entrevoir l'internet sous des angles stratégique et synthétique : même sous forme partielle ou expérimentale, les cartes donnent lieu à beaucoup de commentaires parce qu'elles permettent des représentations globales d'un phénomène appréhendé souvent de manière très fragmentée et aussi parce qu'elles révèlent de nouvelles opportunités de marchés et d'usages.

La cartographie des territoires du réseau d'internet utilise une classe d'objets commune à la représentation de l'espace géographique et à celle de l'espace immatériel d'internet. Elle recouvre plusieurs dimensions. L'une recouvre des aspects topologiques dont on peut faire la représentation, par exemple sous forme de graphe, sans référence à l'espace géographique concret : liens entre serveurs, routes empruntées par les paquets de données, temps d'accès entre serveurs, niveaux de fréquentation de sites etc. Une autre recouvre une matérialité physique de réseau dont on peut assez précisément repérer les dimensions géographiques pour les localiser. L'articulation entre ces deux dimensions est le pari de cette cartographie spécifique et repose sur l'hypothèse que la localisation géographique n'est pas neutre pour l'organisation de l'espace et détermine le contour de nouvelles dynamiques d'entreprises et d'utilisateurs.

Toutefois la simple localisation des éléments physiques du réseau internet sur une carte géographique ne doit pas préfigurer *a priori* de tous leurs services et usages. Il est nécessaire de s'interroger sur la nature et le contenu des objets représentés, et surtout de saisir les usages dont ils sont aussi porteurs en termes de vie sociale et de développement futur pour les territoires : cette cartographie devient doublement pertinente parce qu'elle figure des territoires d'usages, et aussi parce qu'elle est plus explicitement objet d'enjeux. Les cartes ne sont donc pas anodines, elles ne sont pas de simples représentations d'infrastructures.

La cartographie des infrastructures de haut débit en France illustre l'un de ces enjeux. Si l'on peut penser que les processus de déploiement des dispositifs permettant l'accès au haut débit est inégal, encore faut-il tenir compte de l'hétérogénéité du territoire pour rendre compte des réels contours d'une « fracture numérique ». De même les sites web sont concevables comme des serveurs ayant une matérialité physique liée à leur lieu d'hébergement, et une dimension socio-géographique liée à leur commanditaire situé en général à proximité de

l'hébergeur : ils ne peuvent être traités sous le seul angle du nombre et de leur distribution dans l'espace.

Jusqu'à quel point la fracture numérique tient-elle compte de la forme des systèmes de peuplement, des hiérarchies urbaines, des effets d'axes en place? L'analyse de deux grandes régions françaises spatialement bien différenciées, le grand ouest et le sud-est, permet de faire une première évaluation de l'impact des différentes formes de diffusion de l'internet dans les territoires (ADSL, points d'accès publics, centres d'appels, sites web,...).

Mesures spatiales et économiques du web

Marina Duféal

Doctorante
UMR Espace, Université d'Avignon

Marina.dufeal@univ-avignon.fr

En géographie, l'analyse de l'émergence et de la dynamique du web est récente et sort des cadres d'analyse courants, peut être parce que les géographes sont peu enclins à étudier un *cyberespace* décrit par beaucoup comme a-spatial.

Pourtant, l'information transitant sur le réseau lui-même est potentiellement porteuse de sens spatial, et à ce titre mérite de faire l'objet d'une analyse.

Au-delà d'une approche classique de réseaux insistant sur les infrastructures, l'analyse de l'information contenue dans les sites web offre l'opportunité d'une relecture de l'espace, de ses caractéristiques et de ses dynamiques.

Dans le cadre d'un questionnement sur la diffusion spatiale du web, ce sont donc les usages des commanditaires de sites qui attirent l'attention. Ces commanditaires, à l'origine de la création des sites, sont les entreprises, les collectivités territoriales (commune, structure intercommunale, département, région), les établissements scolaires, les universités, les associations...

L'information véhiculée sur les sites renseigne à la fois sur les lieux et les secteurs d'activités les plus réceptifs à l'adoption de l'innovation.

«L'organisation» de ce territoire numérique se compose donc de l'ensemble de ces initiatives qui «projettent» sur le web d'une part l'espace et sa hiérarchisation, d'autre part les activités des lieux.

Ces initiatives sont l'expression de niveaux de développement, de spécialisations économiques, de particularités de l'échange social.

Mais l'étude de la diffusion spatiale des sites web nécessite en amont de constituer une base d'information originale. À partir de nos explorations sur l'annuaire *Yahoo! France*, nous avons constitué une base de données en caractérisant chaque site web par une mesure spatiale et une mesure économique (respectivement à partir du code officiel géographique et de la Nomenclature des Activités Françaises de l'INSEE).

Dans un second temps, nous avons construit quelques indicateurs susceptibles de mesurer la diffusion spatiale du web à l'échelle de l'espace français, et sa diffusion socio-économique à l'échelle du système de villes de l'arc méditerranéen français.

Ce sont les affichages immatériels des villes qui sont étudiés ici, en les articulant aux structures en place : dans le système hiérarchisé de villes d'une part, et dans les compositions économiques urbaines d'autre part.

Usages et comportements d'achat sur internet :

une étude économétrique sur des données françaises

Fabrice Le Guel et Thierry Pénard

CREREG, Rennes 1 et MARSOUIN¹
Faculté de Sciences Économiques, Université de Rennes 1
7 place Hoche - 35 065 Rennes Cedex

(fabrice.le-guel, thierry.penard)@univ-rennes1.fr

Quels sont les facteurs déterminants pour l'adoption et les usages de l'internet ? Afin de répondre à cette question, les enquêtes se sont multipliées ces dernières années tant aux États-Unis qu'en Europe. Souvent, le traitement de ces enquêtes fait appel à des techniques économétriques permettant de bien séparer les effets en jeu et de mesurer précisément leur ampleur. Les modèles de choix discrets de type *Logit* ou *Probit* offrent des techniques économétriques particulièrement adaptées pour étudier des décisions individuelles comme l'acquisition d'un ordinateur, l'abonnement à internet ou l'achat en ligne. Ces méthodes ont été en particulier utilisées dans les études américaines de Brynjolffson et Smith (2001) ou de Goolsbee et Zittrain (1999), toutes deux consacrées aux comportements d'achat en ligne. Notre communication s'inscrit dans la continuité de ces travaux et propose une modélisation appliquée des choix d'adoption de l'internet dans les ménages français et des choix d'achat en ligne. Pour cela, nous disposons d'une enquête effectuée fin 2002 sur les équipements TIC et les usages de l'internet de deux mille ménages en Bretagne. L'intérêt de cette base de données est de pouvoir identifier précisément les facteurs déclenchant l'adoption de l'internet à domicile, ainsi que l'achat en ligne.

Notre étude est originale à plus d'un titre. Tout d'abord, il n'existe pas à notre connaissance d'étude équivalente sur des données françaises. Ensuite, les résultats économétriques obtenus à l'aide de modèles *Logit* sont très significatifs et sont assez conformes avec ceux de Goolsbee et Zittrain (1999). Ainsi, les facteurs explicatifs ne sont pas nécessairement les mêmes pour l'adoption de l'internet et pour l'achat en ligne. De plus, si les facteurs socio-économiques (âge, niveau d'étude, CSP) sont importants, le réseau social des individus joue aussi un rôle clé. Ainsi, un individu aura une probabilité beaucoup plus grande d'acheter sur internet si une large partie de son entourage achète aussi en ligne. Cette idée bien connue des sociologues est donc confirmée en matière de commerce électronique.

Mots clés : diffusion de l'internet, achat en ligne, voisinage, modèle de choix discrets.

¹ MARSOUIN, Môle Armoricaïn de Recherche sur la SOciété de l'information et les Usages d'INternet, <http://www.marsouin.org>.

Questionner l'expression de «fracture numérique»

Éric George

Professeur

Département de communication, Université d'Ottawa
Chercheur, Groupe de recherche interdisciplinaire sur la
communication, l'information et la société (GRICIS)

egeorge@uottawa.ca

L'expression de «fracture numérique» est devenue courante au fil des ans. Celle-ci est employée dans tous les champs, académique, journalistique, économique, politique, etc. Dans le cadre de notre intervention, nous proposons de mettre en évidence la dimension polysémique de celle-ci, en estimant que les enjeux de définitions ont en fait une dimension politique souvent cachée.

Quelques éléments caractéristiques du trafic internet

Philippe Olivier
France Télécom R&D, Issy-Les-Moulineaux

Philippe Owezarski
LAAS-CNRS, Toulouse

Kavé Salamatian
LIP6, Paris

phil.olivier@francetelecom.com, owe@laas.fr, kave.salamatian@lip6.fr

L'internet a connu et continue de connaître une croissance sans précédent tant en termes de nombres d'utilisateurs que d'usages, et finit par avoir une fonction de réseau universel.

D'un point de vue technique, l'internet repose sur le protocole IP qui supporte de multiples applications, dont les diverses caractéristiques se mélangent au sein des ressources du réseau ; les paquets de gestion partagent le même réseau que les paquets de données utiles (il n'existe pas de réseau de signalisation séparé) ; enfin, les mécanismes de contrôle (tels TCP) interagissent fortement avec le trafic offert, produisant un trafic observable grandement modifié. Ces caractéristiques confèrent à la structure du trafic IP un caractère d'une grande complexité, dont une manifestation est par exemple le phénomène de dépendance à long terme ou d'auto-similarité.

La connaissance structurelle de ce trafic nécessite donc d'être étudiée plus ou moins continuellement en fonction de ses évolutions. Techniquement, la nécessité d'effectuer des mesures et des analyses du trafic se justifie plus particulièrement par les besoins suivants :

- Evaluer la demande en trafic des utilisateurs, notamment dans le cadre de la fourniture à venir de classes de service différenciées (COS).
- Dimensionner les ressources du réseau : capacité de traitement des routeurs ; débit de transmission des liens ; taille des buffers aux interfaces. Adapter la gestion opérationnelle de ces ressources à l'évolutivité temporelle de la demande en trafic.
- Contrôler la QoS offerte par le réseau : taux de perte de paquets, délai et gigue de transfert des paquets de bout en bout pour les applications à contraintes temps-réel, débit utile de transport des flux de trafic de données.
- Tester l'adéquation des modèles de performance élaborés au moyen de calculs analytiques ou de simulations, tant au niveau de la validation des hypothèses considérées que de la pertinence des résultats.

Différents types de mesure du trafic peuvent être mises en œuvre : mesures passives ou mesures actives.

Dans le présent état de l'art, nous ne parlerons pas des outils, matériels et logiciels, permettant de mettre en œuvre ces mesures passives et actives. Nous nous attacherons ici à synthétiser les différents modèles de description du trafic qui ont été élaborés au moyen de l'analyse de campagnes de mesures. Celles-ci ne seront que peu évoquées en elles-mêmes. Il s'agira généralement d'observations réalisées à l'aide de sondes passives, plus spécialement adaptées à la fourniture de paramètres descriptifs du trafic. Les sondes actives fournissent quant à elles des informations plutôt liées à la performance du trafic et à la QoS réseau, sujets débordant quelque peu du cadre assigné à cette communication.

TABLE DES COMMUNICATIONS

<i>Métrologie des réseaux de l'internet : principales actions et impact sur les évolutions technologiques</i> Philippe Owezarski	1
<i>Les dispositifs de construction de l'internaute par la mesure d'audience : une approche sociologique</i> Josiane Jouët	3
<i>Application des techniques de fouille de données aux logs : état de l'art sur le Web Usage Mining</i> Doru Tanasa, Brigitte Trousse, Florent Masseglia	5
<i>Adaptive On-Line Page Importance Computation</i> Serge Abiteboul, Mihai Preda, Grégory Cobena	7
<i>Quelques mesures du web comme corpus d'usages linguistiques</i> Russon Wooldridge	9
<i>Analyse sémantique par ordinateur des courriels de la ressource de cybermentorat Academos</i> Jacques Lajoie	11
<i>Le commerce des écritures électroniques</i> Valérie Beaudouin	13
<i>Internet, une cité imparfaite</i> Richard Delmas	15
<i>Les groupes leaders des listes électroniques - Un essai de caractérisation</i> Madeleine Akrich et Cécile Méadel	17
<i>Evolution des métiers de la documentation</i> Alain d'Iribarne, Anne de Fenoyl, Marco Botta, Patrice Hemmel	19
<i>Études des Grilles et des Systèmes Pair-à-Pair : de la mesure au grand instrument d'expérimentation</i> Olivier Richard	21
<i>Mesurer les usages de P2P ?</i> Geneviève Vidal, Vincent Mabillot	23

<i>« Mesurer » des situations de sociabilités médiatisées</i> Paulette Duarte, Mohammed Boubezari, Marie-Christine Couic	25
<i>Invariance d'échelle dans l'internet</i> Patrice Abry, Patrick Flandrin, Nicolas Hohn, Darryl Veitch	27
<i>Topologie d'internet et du web : mesure et modélisation</i> Jean-Loup Guillaume, Matthieu Latapy	29
<i>Écriture, informatique et internet</i> Réal Gingras	31
<i>Ménestrel, un réseau mesuré</i> Christine Ducourtieux	33
<i>Une non-mesure de l'internet : usages subjectifs et rêves d'une bibliothèque numérique, contes et mécomptes d'un bibliothécaire</i> Olivier Bogros	35
<i>Usages des bibliothèques numériques en ligne</i> Houssein Assadi, Thomas Beauvisage	37
<i>La cartographie d'internet, outil de mesure et d'enjeux des dynamiques de territoire</i> Sophie Houzet, Loïc Grasland	39
<i>Mesures spatiales et économiques du web</i> Marina Duféal	41
<i>Usages et comportements d'achat sur internet : une étude économétrique sur des données françaises</i> Fabrice Le Guel, Thierry Pénard	43
<i>Questionner l'expression de « fracture numérique »</i> Éric George	45
<i>Quelques éléments caractéristiques du trafic internet</i> Philippe Olivier, Philippe Owezarski, Kavé Salamatian	47

PROGRAMME

Note : à l'exception des tables rondes, chaque conférence dure 15 à 20 minutes, et prolongée par un débat de même durée.

Lundi 12 mai

Première session : Enjeux scientifiques des mesures

Président : François Baccelli

Horaire	Titre	Auteur(e/s) et orateur/trice	Thématique ou informations complémentaires
9h – 9h20	Introduction	Marc Berthod Suzanne Beauchamp	Adjoint au Directeur Scientifique de l'INRIA Directrice des Opérations, Association les Canadiens en Europe
9h20 – 9h35	Conférence inaugurale : le projet de cette rencontre	Eric Guichard	Responsable scientifique du colloque
9h35 - 9h55	Métrologie des réseaux de l'internet : principales actions et impact sur les évolutions technologiques	Philippe Owezarski	L'état de l'art de la métrologie. Lien entre statistiques enfouies et usages
10h15 – 10h35	Les dispositifs de construction de l'internaute par la mesure d'audience : une approche sociologique	Josiane Jouët	Apports et limites de la mesure d'audience de l'internet
10h55 – 11h10 Pause			
11h10 – 11h30	Application des techniques de fouilles de données aux logs	Doru Tanasa, Florent Masegla , Brigitte Trousse	<i>Le Web Usage Mining</i> : état de l'art
11h50 - 12h10	Adaptative On-Line Page Importance Computation	Serge Abiteboul, Mihai Preda, Gregory Cobena	Mesurer l'importance des pages web. Un algorithme meilleur que celui de Google ?
12h30 – 14h00 Repas			

Seconde session : Faits de langue et d'écriture

Présidente : Émilie Devriendt

14h00 – 14h20	Quelques mesures du web comme corpus d'usages linguistiques	Russon Wooldridge	Mesures des langues vivantes <i>écrites</i> en tirant parti du gigantesque corpus qu'est le web
14h40 – 15h00	Analyse linguistique de la ressource du cybermentorat Academos à partir des échanges électroniques	Jacques Lajoie, Catherine Légaré, Pierre Plante	Analyse sémantique des échanges entre élèves et enseignants
15h20 – 15h40	Le commerce des écritures électroniques	Valérie Beaudoin	Comment l'entrelacement entre publication et conversation sur l'internet réorganise les territoires de l'intellect
16h00 – 16h20 Pause			

Troisième session : Enjeux politiques et professionnels

Président : Frédéric Moatty

16h20 – 16h40	Internet, une cité imparfaite	Richard Delmas	Comment mesures et non-mesures aident l'ICANN à légitimer son monopole
17h – 17h20	Les groupes <i>leaders</i> des listes électroniques. Un essai de caractérisation	Madeleine Akrich, Cécile Meadel	Mesures de l'appropriation de la <i>parole</i> et typologie des profils de locuteurs sur des listes à fort échange
17h40 – 18h00	Evolution des métiers de la documentation en entreprise	Alain d'Iribarne, Anne de Fenoyl , Marco Botta, Patrice Hemmel	Documentation, organisation, technologie et instruments de gestion sociale
18h30 – 19h15	<i>Présentation du Laboratoire des Usages</i>	Claude Guégen	

Le repas du soir est libre.

Pour les personnes intéressées, une autre table ronde informelle peut être organisée de 21h30 à 23h.

Mardi 13 mai

Quatrième session : Réseaux, peer-to-peer et appropriation

Présidents : Yves Lechevallier et Paul Mathias

Horaire	Titre	Auteur(e/s) et orateur/trice	Thématique ou informations complémentaires
9h – 9h20	Etudes des grilles et des systèmes pair-à-pair : de la mesure à la mise en place d'un grand instrument d'expérimentation	Olivier Richard	Des GRID au peer-to-peer : projets actuels, complexité de la mesure des usages
9h40 – 10h00	Mesurer les usages de P2P ?	Geneviève Vidal Vincent Mabilot	Enquête au sein des « professionnels » du peer-to-peer
10h20 – 10h40	Mesurer des situations de sociabilité médiatisées	Paulette Duarte Marie-Christine Couic Mohammed Boubezari	A partir des passionnés de jeux en ligne, l'exemple d'une construction territoriale où s'articulent espaces publics et privés
11h00 – 11h15 Pause			
11h15 – 11h35	Invariance d'échelle dans l'internet	Patrice Abry Patrick Flandrin Nicolas Hohn, Darryl Veitch	Métriques explicatives de l'internet
11h45 – 12h05	Topologie d'internet et du web : mesure et modélisation	Jean-Loup Guillaume Matthieu Latapy	Comprendre les graphes de l'internet et du web, et en retour, imaginer de nouveaux modèles grâce à ces deux exemples
12h25 – 14h00 Repas			

Cinquième session : Pratiques culturelles et pédagogiques

Président : Serge Proulx

14h00 – 14h20	Ecriture, enseignement et internet	Réal Gingras (Prof en ligne)	Une expérience concluante d'enseignement à distance ; mesures et analyses
14h40 – 15h00	Ménestrel, un réseau mesuré	Christine Ducourtieux	Pourquoi un site de médiévistes faisant fonction d'aimant dans la discipline se doit de mesurer son audience ?

15h20 – 15h40 Pause			
15h40 – 16h00	Une non-mesure de l'internet : usages subjectifs et rêvés d'une bibliothèque numérique ; contes et mécomptes d'un bibliothécaire	Olivier Bogros	Une extension de la notion de mesure à partir de la spécificité de la Médiathèque André Malraux de Lisieux
16h20 – 16h40	Usage des bibliothèques numériques en ligne	Houssem Assadi Thomas Beauvisage	Une analyse en profondeur des pratiques des utilisateurs de Gallica (BNF)
17h00 – 17h10 Relâche			
17h10 – 18h30	<i>Table ronde</i>	Présidente : Suzanne Beauchamp	Construire un univers de recherche France-Canada-Europe, de la communication médiatisée par ordinateur à la métrologie des réseaux

18h30 : départ pour la soirée de Gala

Mercredi 14 mai

Sixième session : Territoires de l'Internet

Président : Henri Desbois

Horaire	Titre	Auteur(e/s) et orateur/trice	Thématique ou informations complémentaires
9h30 – 9h50	La cartographie d'internet, outil de mesures et d'enjeux des dynamiques de territoires	Sophie Houzet Loïc Grasland	<i>Le cyberspace</i> , un territoire concrètement construit par les usages
10h10 – 10h30	Analyse spatiale de la diffusion du web dans l'arc méditerranéen français	Marina Duféal	Comment les collectivités territoriales reconstituent <i>pays et terroirs</i> au travers du web
10h50 – 11h10 Pause			
11h10 – 11h30	Usages et comportement d'achat sur internet	Fabrice Le Guel Thierry Pénard Raphaël Suire	Une étude économétrique sur des données françaises : effets du réseau social
11h50 – 12h10	Questionner l'expression de « fracture numérique »	Eric George	Un enjeu politique, bien éloigné des notions de mesure et d'appropriation
12h30 – 14h00 Repas			

Septième session : Mesures d'aujourd'hui et de demain

Président : Éric Guichard

14h00 – 14h20	Quelques éléments caractéristiques du trafic internet	Philippe Olivier Philippe Owezarski Kavé Salamatian	Paquest, flots, protocoles et nouveaux usages
14h40 – 16h00	<i>Table ronde</i>	Président : Loïc Grasland	Cartographie de l'internet, fracture numérique et cybergéographie
16h00 – 16h20 Pause			
16h20 – 17h00	<i>Bilan et programmes futurs</i>		
17h00 – 17h30	<i>Conclusion</i>	Éric Guichard	

